TREATISE ON BASIC PHILOSOPHY

Volume 5

EPISTEMOLOGY AND METHODOLOGY I:
EXPLORING THE WORLD

TREATISE ON BASIC PHILOSOPHY

	1
SEMANTICS I	*Sense and Reference*
	2
SEMANTICS II	*Interpretation and Truth*
	3
ONTOLOGY I	*The Furniture of the World*
	4
ONTOLOGY II	*A World of Systems*
	5
EPISTEMOLOGY & METHODOLOGY I	*Exploring the World*
	6
EPISTEMOLOGY & METHODOLOGY II	*Understanding the World*
	7
EPISTEMOLOGY & METHODOLOGY III	*Philosophy of Science & Technology*
	8
ETHICS	*The Good and the Right*

MARIO BUNGE

Treatise on Basic Philosophy

VOLUME 5

Epistemology & Methodology I:

EXPLORING THE WORLD

D. REIDEL PUBLISHING COMPANY

A MEMBER OF THE KLUWER ACADEMIC PUBLISHERS GROUP

DORDRECHT / BOSTON / LANCASTER

Library of Congress Cataloging in Publication Data

Bunge, Mario Augusto.
Exploring the world.

(Epistemology & methodology ; 1) (Treatise on basic philosophy ; v. 5)
Bibliography: p.
Includes indexes.
1. Knowledge, Theory of. 2. Methodology. I. Title. II. Series: Bunge, Mario Augusto. Epistemology and methodology, 1. III. Series: Bunge, Mario Augusto. Treatise on basic philosophy ; v. 5.
BD161.B86 1983 121 83-3236
ISBN-13: 978-90-277-1523-4 e-ISBN-13: 978-94-009-7027-4
DOI: 10.1007/978-94-009-7027-4

Published by D. Reidel Publishing Company,
P.O. Box 17, 3300 AA Dordrecht, Holland.

Sold and distributed in the U.S.A. and Canada
by Kluwer Academic Publishers
190 Old Derby Street, Hingham, MA 02043, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers Group,
P.O. Box 322, 3300 AH Dordrecht, Holland.

GENERAL PREFACE TO THE *TREATISE*

This volume is part of a comprehensive *Treatise on Basic Philosophy*. The treatise encompasses what the author takes to be the nucleus of contemporary philosophy, namely semantics (theories of meaning and truth), epistemology (theories of knowledge), metaphysics (general theories of the world), and ethics (theories of value and of right action).

Social philosophy, political philosophy, legal philosophy, the philosophy of education, aesthetics, the philosophy of religion and other branches of philosophy have been excluded from the above *quadrivium* either because they have been absorbed by the sciences of man or because they may be regarded as applications of both fundamental philosophy and logic. Nor has logic been included in the *Treatise* although it is as much a part of philosophy as it is of mathematics. The reason for this exclusion is that logic has become a subject so technical that only mathematicians can hope to make original contributions to it. We have just borrowed whatever logic we use.

The philosophy expounded in the *Treatise* is systematic and, to some extent, also exact and scientific. That is, the philosophical theories formulated in these volumes are (*a*) formulated in certain exact (mathematical) languages and (*b*) hoped to be consistent with contemporary science.

Now a word of apology for attempting to build a system of basic philosophy. As we are supposed to live in the age of analysis, it may well be wondered whether there is any room left, except in the cemeteries of ideas, for philosophical syntheses. The author's opinion is that analysis, though necessary, is insufficient—except of course for destruction. The ultimate goal of theoretical research, be it in philosophy, science, or mathematics, is the construction of systems, i.e. theories. Moreover these theories should be articulated into systems rather than being disjoint, let alone mutually at odds.

Once we have got a system we may proceed to taking it apart. First the tree, then the sawdust. And having attained the sawdust stage we should move on to the next, namely the building of further systems. And this for three reasons: because the world itself is systemic, because no idea can

become fully clear unless it is embedded in some system or other, and because sawdust philosophy is rather boring.

The author dedicates this work to his philosophy teacher

Kanenas T. Pota

in gratitude for his advice: "Do your own thing. Your reward will be doing it, your punishment having done it".

CONTENTS OF EPISTEMOLOGY I

PREFACE TO *EPISTEMOLOGY I & II*

We have learned a good deal about many things, we have come to realize that if we value knowledge we should continue to inquire because we shall never know enough, and we keep expanding, applying, and diffusing nearly every bit of knowledge we manage to acquire. So many people are engaged in generating new knowledge, using it, and diffusing it, that economists speak of the *knowledge industry*.

Yet we are still remarkably ignorant about knowledge itself. We still argue whether it is the brain or the mind that knows, and are uncertain about the cognitive abilities of subhuman animals as well as of computers. We have only the most rudimentary information about the development (ontogeny) and evolution (phylogeny) of such abilities, and we continue to ask whether the study of knowledge is a harmless armchair specialty or a useful cross-disciplinary field of inquiry, both factual and theoretical, scientific and philosophical. In short, while the knowledge industry is thriving, the science and philosophy of knowledge are still in bud.

There are several possible explanations for the underdevelopment of our knowledge of knowledge. One could be that epistemology is impossible: this is the view of the radical skeptics and a few others (e.g. Nelson, 1912). Another view is that epistemology and, in general, philosophy has nothing new left to say (e.g. Rorty, 1979).

I find such skepticism and pessimism unacceptable, in view of the rapidly expanding body of interesting epistemological problems raised by science, technology, and the humanities, as well as recent advances in the physiology, psychology, and sociology of perception, learning, and ideation. There must be an alternative explanation for the underdevelopment of epistemology. We may have to blame the conservatism of most philosophers, who stick to the same old problems and rely exclusively on ordinary knowledge as well as on the authority of classical *isms* instead of engaging in a scientific investigation of the problems of cognition and knowledge. If this is the case, then what epistemology needs is not more obituaries but a radical reorientation. I hope this work will contribute to such renewal. This book and its companion, namely Volume 6 of my *Treatise*, concern the basic traits of cognition and knowledge, as well as the

basic principles guiding the acquisition, analysis and utilization of knowledge. Their joint title could well be *Principles of Inquiry*, or *An Inquiry into Inquiry*.

This volume is devoted to general epistemology and methodology; the next, to some epistemological and methodological problems arising in contemporary science and technology. Epistemology, or the theory of knowledge (Fr. *gnoséologie*, Ger. *Erkenntnistheorie*), is the field of research concerned with human knowledge in general—ordinary and scientific, intuitive and formal, pure and action-oriented. And methodology—not to be mistaken for methodics, or a set of methods or techniques—is the discipline that studies the principles of successful inquiry, whether in ordinary life, science, technology, or the humanities.

In this work epistemology is conceived as a merger of philosophy, psychology, and sociology: it describes and analyzes the various facets of human cognitive processes, whether successful or not, and whether or not they bear on everyday matters. Methodology too is descriptive and analytical, but in addition it is prescriptive or normative: it attempts to find out not only how people actually get to know but also how they ought to proceed in order to attain their cognitive goals. Thus both the epistemologist and the methodologist are supposed to describe and analyze experiment, but the methodologist is primarily interested in well designed experiment. In short, whereas epistemology is concerned with inquiry in general, the task of methodology is to find or perfect optimal inquiry strategies.

Yet, for all their differences, epistemology and methodology overlap considerably. Both deal with cognitive processes, so both deserve being called 'epistemology'. After all, it is the same people who succeed and fail in attempting to learn. Both studies should be pursued, and often at the same time, if we are to improve our understanding of the mechanisms of our successes and failures in our cognitive ventures. Fortunately, the two lines of inquiry into inquiry intertwine in the research of a new breed of cognitive psychologists. (See Nisbett and Ross, 1980.)

This book continues an old tradition or, rather, a whole fan of traditions started in ancient Greece and India. But at the same time this work departs from tradition with regard to method. It is hoped that our inquiry into inquiry will be closer to the cognitive sciences and, in general, closer to contemporary research, than to obsolete dogma. More particularly, we shall proceed as follows. We shall pick up the rich legacy of epistemological problems and hints (often optimistically called 'theories') bequeathed to us

by the epistemological tradition. We shall enrich it with some of the problems and findings of contemporary scientific, technological and humanistic research, topping it with new hypotheses compatible with the science of the day—in particular neuroscience, psychology, and social science. And we shall elaborate and systematize the whole with the help of a few modest tools such as the concepts of set and function. However, in contradistinction with the former volumes in this *Treatise*, here we shall adopt a far more modest level of formalization. The result is a greater intelligibility—and length. (The formalizations have been put in parentheses and in the Appendices.)

Finally, we shall try and put our epistemological principles to the test: we shall check whether they account for the actual conduct of inquiry or whether they might help improve it. In other words, we want empirical confirmation for our epistemology, and we want it to be useful to research. And, far from confining our considerations to the physical sciences and technologies, we shall take all the sciences and technologies as our testing ground. This undertaking should prove far more rewarding, though also far harder, than writing obituaries of epistemology in the manner of Wittgenstein or Heidegger.

ACKNOWLEDGMENTS

I am grateful to the 25 generations of students in Argentina, Uruguay, the U.S.A., West Germany, Switzerland, México, and Canada, who took my courses in epistemology and methodology, and asked interesting questions or provided valuable information. I also thank the following persons for their advice, criticism, or encouragement: Joseph Agassi (Philosophy, Tel-Aviv University), Antonio M. Battro (Centro de Investigaciones Filosóficas, Buenos Aires), Hervé Barreau (Fondements des Sciences, Université Louis Pasteur, Strasbourg), Félix Córdoba Alva (Centro de Investigaciones Biológicas, La Paz, BCS, México), Mike Dillinger (Linguistics, McGill), Bernard Dubrovsky (Psychiatry, Physiology and Psychology, McGill), José Ferrater Mora (Philosophy, Bryn Mawr College), Manuel Garrido (Philosophy, Universidad Autónoma de Madrid), Alberto Hidalgo (Philosophy, Universidad de Oviedo), Rolfe Leary (Forest Experiment Station, St. Paul, Min.), James McGilvray (Philosophy, McGill), Claude Panaccio (Philosophy, Université du Québec à Trois Rivières), Michel Paradis (Linguistics, McGill), Rafael Pérez Pascual (Physics, UNAM, México), Miguel A. Quintanilla (Philosophy, Universidad de Salamanca), Osvaldo A. Reig (Estudios Ambientales, Universidad Simón Bolívar, Caracas), Marcel Roche (Estudio de la Ciencia, IVIC, Caracas), William R. Shea (Philosophy, McGill), Jane Stewart (Psychology, Concordia University, Montréal), José Félix Tobar (Philosophy, McGill) and Raimo Tuomela (Philosophy, University of Helsinki). I also appreciate the patience of Jan Hattink and Mrs N. Jones, both with the Reidel Publ. Co. I gratefully acknowledge a research grant from the Social Sciences and Humanities Research Council of Canada. Last, but not least, I thank my housekeeper Mrs Daniela Konieczny.

M. B.
June 1983

SPECIAL SYMBOLS

$\neg p$	*not-p*
$p \,\&\, q$	*p and q* (*conjunction*)
$p \vee q$	*p or q* (*disjunction*)
$p \Rightarrow q$	If p, then q (*implication*)
$p \Leftrightarrow q$	If p, then q, and conversely (*iff*)
$A \vdash B$	Premises A *entail* conclusions B
$=_{df}$	Identical by *definition*
$\{x \mid Fx\}$	The *set* of objects possessing property F
$\varnothing$	The *empty set*
$a \in A$	Individual *a belongs to* set A
$A \subseteq B$	Set A is *included* in set B
$A \cup B$	The set of objects in A *or* in B
$A \cap B$	The set of objects in A *and* in B
$A - B$	The set of objects in A but *not in* B
$A \Delta B$	The set of objects in A *or* in B but *not in both*
$\langle a, b \rangle$	The *ordered pair* of a and b
$A \times B$	The *cartesian product of* A and B, i.e. the set of ordered pairs $\langle a, b \rangle$, where a is in A and b in B
2^A	The *power set* (family of all subsets) of set A
$\lvert A \rvert$	The *cardinality* (numerosity) of set A
$\mathbb{N}$	The set of *natural numbers* (0, 1, 2, ...)
$\mathbb{R}$	The *real line*
$\mathbb{R}^+$	The set of *non-negative real numbers*
$f: A \to B$	The *function* f maps set A into set B
$f(x)$	The *value* of function f at x
$F = \langle F_1, F_2, \ldots, F_n, \ldots \rangle$	A state function for some system
R & D	Research and development
S & T	Science and technology

INTRODUCTION

In this Introduction we shall state the business of both descriptive and normative epistemology, and shall locate them in the map of learning. This must be done because epistemology has been pronounced dead, and methodology nonexisting; and because, when acknowledged at all, they are often misplaced.

1. Descriptive epistemology

The following problems are typical of classical epistemology:

(i) What can we know?
(ii) How do we know?
(iii) What, if anything, does the subject contribute to his knowledge?
(iv) What is truth?
(v) How can we recognize truth?
(vi) What is probable knowledge as opposed to certain knowledge?
(vii) Is there a priori knowledge, and if so of what?
(viii) How are knowledge and action related?
(ix) How are knowledge and language related?
(x) What is the status of concepts and propositions?

In some guise or other all of these problems are still with us. To be sure, if construed as a demand for an inventory of knowledge the first problem is not a philosophical one any more than the question 'What is there?'. But it is a genuine philosophical problem if construed thus: 'What kinds of object are knowable—and which ones are not?' However, it is doubtful that philosophy can offer a correct answer to this problem without the help of science and technology. For example, only these disciplines can tell us whether man can know not only phenomena (appearances) but also noumena (things in themselves or self-existing objects).

Problem (ii) used to be called that of the sources of knowledge: do we know by revelation, intuition, sense experience, action, or reasoning? It

may now be reformulated as the problem of cognitive mechanisms, and it is of interest to psychology, not only to philosophy.

Problem (iii)—about the possible contribution of the subject—used to trigger automatic responses, in particular 'Nothing' (empiricism) and 'Everything' (rationalism). It can now be investigated with the help of psychology.

Problem (iv) is ill-conceived, for it presupposes that there is a single kind of truth. Actually there are at least four kinds of truth (or four kinds of proposition to which a truth value can be assigned): logical, mathematical, factual, and philosophical. Logical and mathematical truth have become the property of logic and metamathematics. The problem of philosophical truth belongs to both descriptive and normative epistemology. And truth of fact, the concept occurring in all inquiries into matters of fact, is investigated by semantics—and perhaps can also be approached by psychobiology.

Problem (v) is that of truth tests and criteria, and is therefore best left to normative epistemology or methodology.

Problem (vi) is that of the fallibility of factual knowledge—or, to put it in positive terms, of the corrigibility of hypotheses and data. And it may or may not be linked to the probability calculus. (It is not so linked in the present work.)

Problem (vii) is that of the scope of pure reason, the nature of formal science, and the existence of (true) a priori statements of fact—hence it is actually a whole family of problems.

Problem (viii) is central to the philosophy of technology as well as to dialectical materialism and pragmatism. Any examination of it should include an analysis of the role of knowledge in the design and implementation of plans, as well as a discussion of the value of praxis in posing cognitive problems or even in evaluating truth claims.

Problem (ix) is at the very core of the philosophies of language and mind. However, it should be handled with an eye on developmental psychology, neurolinguistics, and psycholinguistics.

Finally problem (x) belongs also to semantics and ontology, as well as to the philosophy of logic and mathematics.

Note that all of the above problems continue to draw the attention of some people and may thus be said (metaphorically) to be alive. (A problem is "dead" if no living being deals with it.) And they are "alive" because they continue to be open, i.e. not fully solved. Note also that epistemology has ceased to hold a monopoly on those problems: it shares them with other

disciplines, be they other branches of philosophy or special sciences. In other words epistemology is (if not *de facto* at least *de jure*) no longer an independent discipline. The moral is clear: Do not pretend that epistemological problems can be attacked without the help of allied disciplines, in particular (though not exclusively) semantics, psychology and social science.

Nor are the above the only epistemological problems left, not even the only classical ones. Here is a short and haphazard list of problems on which epistemologists and sociologists of knowledge are currently working:

(xi) What is the role of mathematics in the generation of factual knowledge?
(xii) What is rationality?
(xiii) Can belief and plausible reasoning be regimented or formalized?
(xiv) What are problems, methods, approaches, hypotheses, theories, and rules?
(xv) What are the peculiarities of tacit knowledge (*know how*) and of explicit knowledge (*know that*)?
(xvi) What are the commonalities and peculiarities of description, classification, explanation, and prediction?
(xvii) What are basic science, applied science, and technology?
(xviii) What, if any, is the role of morality in inquiry?
(xix) How does the social matrix influence the cognitive activities?
(xx) What are the characteristics of learning (or inquiring) communities?

Note that, whereas traditional epistemology focused on knowledge, modern epistemology is also interested in the applications of knowledge. Second, whereas classical epistemology was concerned solely with the knowing subject and his (never her) accomplishments, or both, modern epistemology is also interested in learning communities as well as in the social function of cognition and in the social constraints on it. Consequently some of its problems are shared by social science, in particular the sociology of knowledge and social psychology. Recalling our comments on some of the typical problems of traditional epistemology, we must conclude that epistemology is only one of the cognitive sciences and it has a considerable overlap with its sister disciplines, in particular psychology (cognitive, developmental and evolutionary) and sociology (in particular social psychology and the sociology of knowledge).

2. Normative epistemology (methodology)

Unlike descriptive epistemology its normative partner, methodology, has only weak roots in Antiquity and the Middle Ages. Leaving logic aside, they amount to a handful of rules, such as Plato's injunction to shun opinion (*doxa*) and seek only certain knowledge (*episteme*), Aristotle's practice of defining everything, Hippocrates' recommendation to abstain from supernaturalistic explanation, and Ptolemy's advice that astronomy be restricted to describing phenomena (appearances).

The idea of a universal method for acquiring knowledge did not appear until *ca.* 1600. Discoursing on method became then fashionable, even a craze. There was hardly a scholar who doubted that, if only the right method could be found, mankind would attain instant wisdom. Bacon swore by induction and Descartes by conceptual analysis and deduction from indubitable first principles. Leibniz dreamed not only of a logical algorithm (*characteristica universalis*) but also of an *ars inveniendi* leading securely to first principles. It was a time of mythodology and methodolatry rather than methodology. The illusion lasted nearly three centuries.

Methodology proper did not come until the first modern philosophers of science, in particular Whewell (1847). Although some investigators deny the existence of methodology, it is easy to find methodological problems, many of which have not yet been given satisfactory solutions, or at least generally accepted ones. Here is a sample of rather classical problems of this kind.

(i) Is there a single best way of advancing knowledge and, if so, which is it?
(ii) What is the scientific method?
(iii) What kinds of convention do we stipulate and which are their roles in inquiry?
(iv) What are the rules of correct definition? And are there creative definitions?
(v) What are the principles of correct classification?
(vi) What is the role of regulative (or heuristic) principles in inquiry?
(vii) What kinds of theory should we prefer: descriptive (phenomenological) or explanatory (mechanismic), deterministic or probabilistic, single-level or multilevel?
(viii) What is the point of axiomatics?
(ix) What is testability: confirmability, refutability, or either?

(x) If theories are tested as wholes, how can we spot their false components?

All of these problems are still topical and some of them are still open. Whereas some of them, in particular the first two, are discussed even in bull sessions, others are the subject of technical papers or books. But all of them are likely to be faced, sooner or later, by all investigators. For example (v), or Aristotle's problem, is very much alive among biologists, particularly in connection with numerical and evolutionary taxonomy. Problem (vi), Kant's, comes up in every discussion of correspondence principles. Problem (vii) is confronted by anyone who proposes to build a factual theory or model. Problem (viii), Hilbert's, is discussed by both friends and foes of axiomatics. Problem (ix) is the bone of contention between confirmationists and refutationists. And Duhem's problem (x) is faced by any investigator in the presence of some evidence against an otherwise well confirmed theory.

Here is a random sample of methodological problems that have come more recently to the fore:

(xi) Are properties dispensable? If not, how are they to be represented?
(xii) Can there be any techniques of theory construction?
(xiii) What are the advantages of the reduction (deduction) and of the merger (synthesis) of theories and entire disciplines?
(xiv) How should hypotheses and theories be evaluated? In particular, is empirical confirmation necessary and sufficient?
(xv) Do we have to test all the alternative hypotheses that might account for a given group of facts, or only those enjoying some empirical or theoretical support?
(xvi) How are revolutionary new ideas or procedures to be evaluated? Do they have to be compatible with some of our previous knowledge or can they be totally deviant?
(xvii) What would be the value of measurement and experiment if the experimenter had powers of psychokinesis and precognition?
(xviii) What can be the value of empirical tests if they are designed with the help of theories?
(xix) How should methods be tested?
(xx) How should the rules of statistical testing (e.g. for significance or for randomness) be tested?

Problem (xi), prompted by the nominalist mistrust of properties and their conceptualization (attributes), is both philosophically and technically interesting. Problem (xii) was initially raised by the believers in the possibility of fashioning recipes for inventing, and has recently been revived by some computer enthusiasts. (If empiricists are right in holding that theories are data packages or distillates, why not build or program an electronic theorist?) Problem (xiii) comprises the questions of reductionism and of the unity of science. Problem (xiv) questions the empiricist tenet that empirical confirmation is not only necessary but also sufficient to canonize hypotheses and theories. Problem (xv) questions the refutationist conviction that experimental scientists have all the time in the world to try and refute any hypotheses or theories that may occur to us. Problem (xvi) is that of weighing scientific revolutions and distinguishing them from crackpot proposals. Problem (xvii) suggests that experimental science presupposes the falsity of parapsychology—which in turn poses the problem of the empirical testability of belief in the paranormal. Problem (xviii) amounts to this: Are empirical procedures contaminated by theory to the point that they have no test power? And the last two problems pose the difficult question of evaluating the goodness of methods and rules.

Let the above sample suffice to show that methodology can be neglected only at the risk of making bad mistakes. And note the salient characteristic of methodological problems in contrast to problems in descriptive epistemology (Section 1). Whereas the latter are questions of fact—How do we know *X*?, What is cognitive item *Y*?, and the like—methodology handles typically value problems—What is *X* worth?, What is the correct way of doing *Y*?, and the like. Consequently, whereas an epistemological problem is answered by a description, a definition, or a theory, a methodological problem is answered, in the last analysis, by a value judgment or by a rule of procedure—either of which must in turn be justified in terms of theory, experiment, or praxis. In short, methodology, like logic and decision theory, is a normative discipline.

3. Epistemology and Biology

All cognitive activities, from sniffing and exploring to theorizing and forecasting, are biological functions—as well as social ones in the case of social animals. They are all aspects of the adaptation of animals to their environment as well as of their activity in altering their environments to suit

their needs. Hence biology is, or ought to be, interested in cognition, and all the cognitive sciences are, or ought to be, based on biology.

The view that cognition is a biological process emerged in the late 19th century, which saw the beginnings of experimental psychology as well as of comparative (or evolutionary) psychology, later to be supplemented by developmental (or genetic) psychology. Perception ceased to be a subject of philosophical speculation to become a problem for physiologists and experimental psychologists. Even problem solving and other higher mental functions became eventually the object of experimental investigation and, more timidly, mathematical modelling. Darwin, Helmholtz, Mach, Lloyd Morgan, Yerkes, Köhler, Piaget, Békesy and Hebb were among the pioneers of the biological approach to psychology and, in particular, the study of cognitive abilities. Today this study is pursued vigorously by physiological psychology (cf. Hebb, 1966; Milner, 1970; Bindra, 1976). This approach to cognition must not be confused with *cognitivism*, or the information processing view, just as the study of behavior must not be identified with behaviorism. Indeed the gist of cognitivism is that biology is irrelevant to the understanding of cognition, whereas computer science is essential to it (cf. Simon, 1979; Pylyshyn, 1980).

The *rapprochement* between psychology and epistemology, on the one hand, and biology on the other, was not limited to the scientific community but fascinated also some philosophers. Thus Nietzsche, the real founder of pragmatism, held that lies may be just as useful for life as truths, and claimed that the sole justification for research lies in its usefulness to preserve life. Likewise Spencer is the author of the famous dictum "Science is for life, not life for science". (Many contemporary statesmen and bureaucrats hold the same narrow-minded view of science without having read either Nietszche or Spencer: we keep reinventing falsity.) Spencer also wrote the celebrated sentence echoed in our days by Konrad Lorenz: "What is *a priori* for an individual is *a posteriori* for a species". (That is, the individual does not have to start from scratch, for it inherits a number of traits won in the course of evolution.) And Mach—noted for his important discovery that sensation is a function not only of stimulation but also of inhibition—wrote occasionally of the "transformations and adaptations" of thought. He also wrote about "the adaptation of thought to sensation" (which he regarded as the sole firm source of knowledge), as well as about the conflict between the two (for him the sole source of problems). But he did not spell out how cognitive behavior can be beneficial or harmful, let alone how selection could uproot incorrect cognitive habits.

Neither Nietzsche nor Spencer nor Mach may be said to have built a biological psychology, let alone an evolutionary one. (See Čapek (1968) for the opposite view.) The same holds for Peirce, another thinker influenced by Darwin's theory, and who stated that the truest beliefs are those that "live" longest—an idea refuted by the persistence of superstition. Nor is evolutionary epistemology related to Popper's analogy between natural selection and refutation, let alone to his speculations on disembodied "world 3" items (Popper, 1972). Indeed the theory of evolution applies only to concrete systems, namely biopopulations, not to disembodied objects such as ideas in themselves (Bunge, 1981a).

If epistemology is to become scientific it will have to learn not only from neurophysiology and physiological psychology but also from developmental (genetic) psychology and from whatever evolutionary biology has to say about cognition. That is, it will have to include evolutionary epistemology. So far this is little else than a thesis and a program. The thesis is this: "Our cognitive apparatus is a result of evolution. The subjective cognitive structures are adapted to the world because they have evolved, in the course of evolution, in adaptation to that world. And they match (partially) the real structures because only such matching has made such survival possible" (Vollmer, 1975, p. 102). This is a programmatic thesis: it orients a whole research project that has still to be implemented not only with more observations in comparative psychology and ethology but also with theoretical synthesis and speculation, as well as the conceptual clarification (of, e.g. the hazy notions of cognitive structure and real structure). However immature, the project of evolutionary epistemology is well grounded, not wild: it fits in with evolutionary biology.

4. Epistemology and Psychology

Except for physiological psychologists and the rare zoologist, ethologist, or neuroscientist, cognition is usually regarded as a suprabiological phenomenon and studied separately from the nervous system. One main reason for this non-biological approach is the persistence of the ancient myth of the immaterial mind. This myth is not only a component of most religions and popular world views but also of most psychological doctrines (in particular psychoanalysis) and most philosophies of mind. (For a defense see Popper and Eccles (1977), and Eccles (1980); for criticisms see Bindra (1980), Bunge (1980a, 1981a), and Hebb (1980).)

The detachment of cognition from its organ, namely the nervous system, deprives psychology and epistemology of substance and depth: it is like discoursing about the weather as if it were something different from the state of the atmosphere. By the same token it prevents psychology and epistemology from benefiting from the findings of neuroscience and, in general, it isolates them from all the other cognitive fields. (Isolation is characteristic of protoscience and pseudoscience.)

If mind and matter are distinct substances, then it is hard to understand how the former can get to know the latter, except by adopting some form of idealism. In this case matter does not exist except as a figment of our own mind, and the acquiring of knowledge about anything is reduced to the mind knowing itself or other minds, whether individual or supraindividual. This may work for logic and mathematics but it fails to account for natural science and technology, as in these fields one must take the real world for granted if one wishes to learn something about it. In fact, there would be no point in exploring the world unless it were out there all by itself. In particular we would waste no time experimenting.

On the other hand materialism has no trouble understanding, at least in principle, how a material system (the brain) can get to know something about other material things or even about itself. Indeed, according to materialism (and physiological psychology) every cognitive process is a string of events in some central nervous system. So it may be imagined that it can match other material processes, much as an electrocardiograph transduces voltage variations into graphs, or a radio receiver transforms electromagnetic waves into sound waves. Moreover, materialism happens to be the ontology of modern science, which, far from countenancing disembodied or ghostly entities, assumes that all existents are lawfully changing concrete (or material) things, though not necessarily of the same kind. (See *Treatise*, Vols. 3 and 4.) For these reasons we choose materialism, and more particularly scientific materialism (Bunge, 1981a), as the general ontological view which our epistemology is to fit.

This approach to the study of cognition contrasts with both traditional mentalism and neomentalism, or the philosophy of mind adopted by most workers in the so-called *cognitive science*. This is the collective name of three disciplines concerned with cognition: cognitive psychology, linguistics, and artificial intelligence. Whatever the accomplishments of these fields, they share an important shortcoming, namely, that all three ignore the nervous system and are therefore isolated from neuroscience and psychobiology. Furthermore, cognitivism has produced no new theoretical

models and it has ignored motivation and response production, so it cannot possibly offer a comprehensive framework for psychology (Bindra, 1982). In particular, talk about human information processing and computing is just a metaphor and, as such, no substitute for a theory. (Recall Hobbes' characterization of metaphor as *ignis fatuus*.)

Nor is our approach in sympathy with behavioristic psychology. The investigation of cognitive processes includes the observation and measurement of behavior but cannot be restricted to it because the motivations, intentions, cogitations, goals and plans of the inquiring subjects are not observable. Besides, what the inquirers do is totally unintelligible to an observer who does not have a similar background. For example, a chemist, a chemical engineer and an alchemist may wear white laboratory coats, handle roughly the same instruments and the same chemicals, and even go through similar motions, while doing entirely different things because they think differently and have different goals and methods. For this reason the fashionable anthropological and sociological studies of the scientific community (e.g. Latour and Woolgar, 1979), when limited to observing the behavior of scientists, are useless when not positively damaging to the science of science.

5. The place of epistemology

Many disciplines besides descriptive and normative epistemology study cognition and its outcome, i.e. knowledge. Zoology and ethology are interested in cognition; so are neuroscience and psychobiology (or biopsychology), as well as the non-biological approaches to cognition, in particular neomentalism, linguistics, and artificial intelligence. Nor are these the only members of the family that could bear the hybrid name *cognitology*. All of the social sciences are interested in cognition and knowledge: in the former because learning is largely a social process, in the latter because knowledge is a social force—or, rather, people who know something out of the ordinary can affect the behavior of others. In particular, anthropologists study the functions of the learning professionals, be they shamans or molecular biologists; archaeologists conjecture what ancient peoples must have known in order to manufacture the artifacts that are dug out; social psychologists investigate the social constraints on cognition as well as the effects of knowledge on social structure; and historians—particularly historians of science, technology,

or the humanities—are interested in the traditions and revolutions in modes of inquiry and belief, as well as their relations with environmental, demographic, economic and political factors.

All the members of cognitology have a common referent, namely the inquiring system (individual or group). Therefore there is—inevitably and happily—some overlap between the various sciences of cognition and knowledge. However, a common reference does not ensure identity, for one and the same subject matter can be studied from different viewpoints—i.e. one can ask questions of different kinds about one and the same entity. This holds, in particular, for the scientific and the philosophical approaches to cognition and knowledge. For example, whereas the scientist may study perception and perceptual illusion as sources of knowledge and error respectively, the philosopher may also study the scope of perception, the nature of perceptual knowledge and its differences from conceptual knowledge, as well as the contrast between appearances (as presented in perception) and reality (as conjectured by theory). Whereas the scientist may be interested in the way children come to know about objective constancies (e.g. conservation laws), the philosopher may puzzle the nature of law statements, their relation to objective patterns, and the role of such statements in science and technology. And whereas the scientist may investigate the origin—psychological or historical—of theories, the philosopher may study the very nature and role of finished theories, as well as the conceptual (rather than psychological or cultural) differences between them.

Both approaches to cognition and knowledge, the scientific and the philosophical, are necessary. The scientist could not even start studying, say, the process of hypothesis formation (in some brain and under certain social circumstances) unless he had a reasonably precise and correct idea of what a hypothesis is—which idea it behoves philosophers to exactify. And the philosopher should be happy to learn that somebody studies the genesis, evolution, and demise of hypotheses, not only in highly educated human adults but also in children, primitive man, and even in some subhuman species.

Cognition and knowledge can then, nay must, be studied *both* scientifically and philosophically. The two approaches are needed because they are mutually complementary and they stimulate and control one another. (Only bad science is the enemy of good philosophy, and bad philosophy the rival of good science.) Moreover there is no clear demarcation between scientific and philosophical epistemology, and none should be invented.

(Likewise there is no borderline between science and ontology: see *Treatise*, Vol. 3, Introduction.) This way of conceiving of epistemology is likely to advance our knowledge of cognition and knowledge, for each partial view can be allowed to enrich and check the other.

An additional advantage of conceiving of philosophical epistemology as a member of the large family of cognitive sciences (or *cognitology*) is this. Ordinarily epistemology is regarded as being far removed from both ontology and ethics. In our perspective cognition is only a special kind of biological process, and therefore it is an object of study of the ontology of organisms; and organisms capable of sharing knowledge are members of some society or other, and therefore the object of study of the ontology of society. Thus any reasonably true generalizations about cognition may be placed in the intersection of the science and the philosophy of cognition.

Nor is epistemology alien to ethics or the study of morality. Indeed normative epistemology is about rules of correct (or successful) inquiry; and rules, unlike laws of nature, can be either obeyed or broken with honest or dishonest intentions. Besides, all knowledge is valuable in one way or other; and anything valuable can be well used or ill used, shared or stolen—as well as traded and faked. The pursuit of knowledge must then involve a moral code. In fact it does involve one: recall the commandments to prefer truth to error, to check conjectures, not to make up data, to seek criticism, to acknowledge truth in the rival and error in the friend, to share information, and so on. This code of inquiry intersects with methodology: every correct procedure of inquiry is intellectually honest even though it may be morally objectionable on other grounds.

What about logic and mathematics in their relation to epistemology? These two formal sciences provide the backbone of all the other sciences but they do not investigate cognition and they are not in a position to discriminate factual truth from factual error. In short, they are not members of cognitology. To be sure, logic used to be regarded as the *organon* of inquiry and thus as identical with methodology. But we have learned since the beginning of the modern period that logic may at most discover (logical) error: it is not a guide of inquiry but a constraint on it. We now tend to conceive of logic as the intersection of mathematics and philosophy, as well as a tool of conceptual analysis and criticism in all cognitive fields.

In conclusion, epistemology is not an independent discipline. It must make use of logic and can make use of mathematics, and it belongs to the family of cognitive sciences. Moreover it has considerable overlaps with all

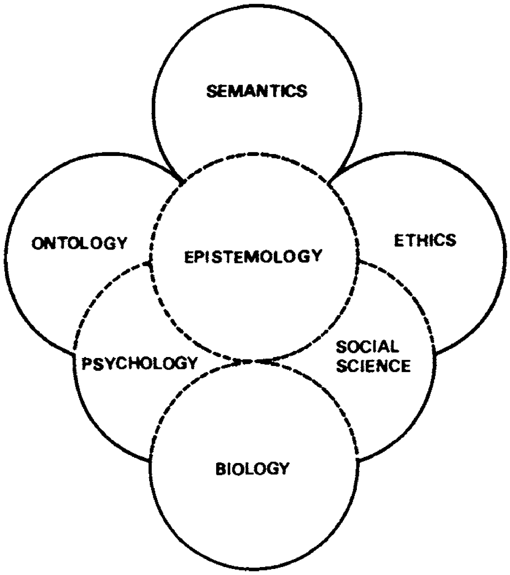


Fig. 0.1. The place of epistemology among the cognitive and philosophical sciences. Note the overlappings with neighbouring fields and the lack of clear cut boundaries between them.

of these disciplines as well as with ontology and ethics, not to mention semantics, from which it is often indistinguishable. See Figure 0.1.

6. Uses of epistemology

There should be no need to justify research on cognition and knowledge: only barbarians or yahoos would question the worth of inquiring for the sake of knowing. But since scientists and philosophers are besieged by barbarians and yahoos, it won't hurt to argue for the practical uses of such studies.

To begin with, learning theory, though still in its infancy, has already had a profound impact on the theory and practice of teaching. (See Suppes (Ed.), 1978.) Secondly, behavior therapists are having some success in correcting bad habits, in particular bad cognitive habits, on the basis of the hypothesis that whatever has been learned can be unlearned. (See Bandura, 1974; Wolpe, 1976, 1978.) Thirdly, neurologists have started to cope with learning disabilities attributable to congenital neural defects or brain injuries. (See Hécaen and Albert, 1978; Pincus and Tucker, 1978.) In short

the psychology of cognition, though still very young, has proved its practical usefulness. Much more can be expected from it if it sheds the myth of the immaterial mind as well as the superficiality and dogmatism of behaviorism.

The social studies of cognition and knowledge have made important inroads as well (cf. Merton, 1973). The most important of them has been showing (not just asserting) that social circumstances condition our mode of perceiving, thinking, and evaluating, so that cognition and knowledge cannot be adequately understood except in their social context. Moreover it has now become clear that, at least in man, knowledge and sociality are coextensive: neither exists without the other. This discovery has had a profound impact on the history of science and technology, which until relatively recently followed the pattern of the history of ideas, investigating ideas in a social vacuum. (Inevitably, the discovery of the social dimension of cognition has led to sociologism, or the neglect of the brain as well as individual circumstances.)

Philosophical epistemology is a little harder to defend nowadays. However, if we care for science, technology, and the humanities, we should also care for their epistemological presuppositions—such as that there is an independent reality and that it can be known if only in part. We should dig up, cleanse, analyze, and systematize such principles. (For a short list see Ch. 15, Section 3.1.) It would seem obvious that, the better we know how we can get to know, the better we can improve (or block) the learning process, particularly in science, technology, and the humanities. So, a study of the epistemological presuppositions of research, as well as of any other tacit assumptions of scientific and technological research, should pay off in practical results.

As for normative epistemology, it is often claimed that it is no more needed for doing science or technology than a knowledge of grammar is needed for writing well or of logic for thinking correctly. True, some people untutored in logic and grammar think and write well, and most successful scientists and technologists have not been exposed to methodology. However, it is also true that the study of grammar improves good writing, that of logic good thinking, and that of methodology good research. Most people untutored in logic and methodology are prone to incur fallacies and find it difficult to design research strategies even for solving simple problems. In his classic *Thinking* (1958) Bartlett concluded that "especially with naive, although intelligent searchers, it is most likely that the method of search adopted will be an uneconomical one".

However, most scientists and technologists are indifferent to, when not contemptuous of, methodology. Philosophers tend to impute this attitude to excessive specialization, lack of philosophical curiosity, or even illiteracy. Scientists and technologists feel on the other hand that methodologists misrepresent what scientists and technologists do or ought to do (e.g. Medawar, 1969). Actually the situation is much worse than either party thinks, for *both* are right. Indeed there seem to be few if any persons capable of giving a truthful and balanced account of the actual conduct of inquiry—even of their own. Author *A* believes only in logic, *B* only in intuition; *C* swears by induction and *D* by deduction; *E* by experiment and *F* by theory; author *G* is a gradualist and *H* a catastrophist with regard to the evolution of knowledge; thinker *J* ignores society altogether and *K* places knowledge in society—and so on and so forth. Every one of them can cite examples in his favor, so neither of them is totally wrong. But none of them has built a system accommodating the many and apparently conflicting features of inquiry: each of them offers a one-sided view, and each is anxious to have it accepted as his own original *ism*.

We must try and strike a golden mean between *methodolatry* and *methodoclasm*. The former is the view that method is more important than product. (Paradigm: the Chinese dictum "Do not give me a fish: instead, teach me how to fish.") Extreme methodolatry will promote the teaching of methods separate from subject matter, and so the student will learn neither. On the other hand methodoclasm holds that only knowledge is important, not how we come to know. (Paradigm: "What matters with a piece of news is whether it is true or false, not the means—hearsay, reading, intuition, etc.—we come by it.") Extreme methodoclasm will promote teaching the results of research with neglect of the way research was initiated, planned, conducted, and evaluated.

A balanced view will hold that both the process of inquiry and its product are worthy of study. This view has two advantages. One is that it allows one to examine and evaluate methods in the light of results and conversely—the virtuous circle principle. (See Bunge, 1962; Rescher, 1977.) Another advantage is that a study of methodology serves to rectify a number of misconceptions about inquiry. Some scientists have been known to be busy on problems other than those they believed they were investigating; at a loss to formulate the most obvious hypotheses or even reluctant to make any conjectures; incapable of clarifying key notions they used daily; totally uninterested in neighboring fields; ignorant of the ultimate goals of scientific research; contemptuous of theory or, more

rarely, of empirical test; believing that all the mathematics they need is a handful of statistical recipes—or, on the contrary, that mathematics suffices for science. Any such scientist would profit by some methodology. And even the most experienced and successful researchers can benefit from pondering, once in a while, over such rules as: "Place your problem in a wider context", "Look for deeper (non-apparent) variables", and "Do not underrate chance". (See Ch. 15, Section 3.2, for a short list of regulative principles of research.)

7. Concluding remarks

Epistemology teems with open problems: a sign that it has a right to life. But epistemology is still immature: a sign that it has a long way to go. It cannot go it alone and in fact it has interacted vigorously with natural science, in particular physics, since the 17th century. But until recently epistemology has had only very weak ties with evolutionary biology, neuropsychology, and social science—not surprisingly, for these came long after epistemology was born.

Epistemology must mesh in with biology, psychology, and social science, because the learning process happens to occur in and among animals. The immaterial and isolated (and male and adult) knowing subject of traditional epistemology must be replaced with the inquiring brain, or team of brains, embedded in society. The unchanging mature mind must be replaced with developing and evolving animals exploring their environment as well as themselves, and changing both with the help of the knowledge gained in such inquiries. Epistemology, in sum, must be biologized and sociologized.

However, this proposed reconstruction of epistemology with the help of biology, psychology, and social science, should not be taken as a pretext for neglecting the epistemological problems that can be handled independently. Thus regardless of whether or not conjecturing and jumping to conclusions are inborn human dispositions; regardless of the particular neural systems that do the conjecturing and the jumping to conclusions; and regardless of the way social circumstances stimulate or inhibit such brain processes, hypotheses and inferences have certain logical, semantical, and methodological properties (e.g. generality, factual reference, and testability) that can and must be studied separately from the corresponding concrete biological and social processes. Likewise we do not need biology,

psychology or social science to check whether a computation is correct or whether the components of an electric circuit have been assembled correctly.

In short, up to a certain point methodological inquiry can be pursued independently of descriptive epistemology. This relative independence explains why we have been able to learn (and stipulate) something about the *correct* way of going about forming and combining concepts and propositions, classing, comparing, analyzing, etc. without really knowing much about the neurophysiology (or even the global psychology) of such operations. Nevertheless this situation is anomalous and reminds one of the engineering feats of the ancients, which had no scientific basis, rather than of modern technology, which is science-based. We may conjecture that a more intense and deeper scientific investigation of cognition, and in particular of the neurophysiology of thinking, may bring about a revolution in methodology or at least a more intimate understanding of the relation between the *is* and the *ought* of knowledge. But for the time being we are faced with the relative independence and more advanced state of development of normative epistemology. This explains why it gets the lion's share in this book.

PART I

COGNITION AND COMMUNICATION

CHAPTER 1

COGNITION

This book is about knowing and knowledge, error and ignorance: about the cognitive process, its pitfalls, and its outcome. When one perceives or conceives an object, investigates it, manipulates it or evaluates it, one engages in a cognitive process. With some ability and luck—i.e. under suitable circumstances—one will learn something about the object of inquiry. Whatever one has learned in this process, added to what he may have known beforehand, is one's *knowledge* of the object.

The distinction between a cognitive process—in particular an inquiry—and its result at a given moment is an instance of the process-state distinction. (A process may be construed as a sequence of states: Vol. 3, Ch. 5, Section 2.3.1.) It is a useful distinction because, for certain purposes, the end result is of little interest, whereas for others it is all that matters. Thus the neuroscientist and the psychologist are usually more interested in the learning process—and its disruptions—than in its accomplishments, whereas the latter attract most of the attention of traditional philosophers as well as of historians. Here we shall be concerned with both.

We ask indeed *Who* (or *what*) *can know what, and how*? This complex question summarizes the problematics of all the disciplines—philosophical, scientific and technological—concerned with cognition or with knowledge. The first part of the question, namely *who* (or *what*) knows, is handled by psychology (both animal and human) as well as by the sociology and history of knowledge. The second part, i.e. *what* can get known, interests epistemology and the history of science, technology, and the humanities. And the third part, namely *how* (by what means, mechanisms, or methods) we may get to know, is addressed by all the cognitive sciences, particularly methodology.

There are three fruitful approaches to the study of cognition: the neurobiological, the psychological, and the philosophical ones. All three are concerned, each in its own way, with the same problematics: internal representation, imagery, conception, conjecturing, systematization, inference, problem solving, criticism, and their kin. The neurobiological study of cognition attempts to lay bare the neural mechanisms of cognitive processes. The psychological study tries to describe them at the molar or

phenomenological level, i.e. without inquiring into their neurophysiology. And the philosophical (epistemological) study endeavors to elucidate those concepts, as well as to discover the contribution of the various modes of knowing to the common enterprise, which is the perfectible knowledge of the world, including ourselves and our cognitive abilities.

Of these three approaches the oldest is the philosophical one, and it is fair to say that philosophers have so far contributed no less to the science of cognition than either psychologists or neuroscientists. But it is also fair to say that the philosophy of cognition is now at the end of its tether: one can only grasp so much of reality without the help of experiment and modeling. Psychology took up the problem of cognition ever since it became an independent science. True, neither behaviorists nor psychoanalysts placed much value on investigating cognition, and both schools ignored that which does the knowing, namely the central nervous system. But other schools and fields, in particular primatology, comparative psychology, developmental psychology, and the Gestalt group, continued to work on cognition.

Interest in cognition has been stronger than ever since about 1960, both among psychologists and neuroscientists. However, many of the former have adopted a new form of mentalism, where analogies with artificial information processors are central whereas the nervous system is all but ignored. They hold cognition to be a matter of software (mind) not hardware (brain). We shall ignore this new version of the old doctrine of the soul and shall adopt instead the psychobiological approach sketched in Vol. 4, Ch. 4. (See also Bindra, 1976; Bunge, 1980a; Hebb, 1980.)

According to psychobiology (or neuropsychology) every mental function is a brain function, but the converse is false. (There are "household" functions, such as metabolism and the synthesis of proteins and hormones, which are nonmental.) This view unites psychology with biology and suggests a vast research project, both experimental and theoretical, that is only beginning to explain the mind and, in particular, cognition. (The project is not new, for it is found already in Henry Maudsley's *The Physiology of Mind* (1876), which gained a considerable readership but few followers.) Because this project is only in its initial stage, many of its hypotheses and theories are speculative, and it offers many more problems than solutions—a sure sign of youth in science. So, it should be judged not so much for its accomplishments as by its promise and philosophical soundness, as well as by the failure of its alternatives, namely mentalism and behaviorism.

1. The knowing subject

1.1 *The Cognitive Organ*

Looking for external objects and perceiving some of them (or forming the idea that we have failed to do so); watching for internal bodily signs and feeling them (or realizing that we failed to perceive them); learning motor skills and intellectual ones; forming concepts or propositions, questions or commands; identifying and classing, inferring and computing, criticizing and theorizing; puzzling and pondering, expecting and planning; speaking and listening to talk—all these and many more are *cognitive* acts or processes. They all generate, transform or utilize bits of knowledge.

Our initial assumption is that *every cognitive act is a process in some nervous system*, whether human or not. To put it negatively: *There is no knowledge in itself*, i.e. separate from the cognitive processes occurring in some nervous system or other. In particular there are no ideas except in brains of certain kinds—presumably those of mammals and birds. No autonomous ideas, i.e. ideas detached from brains, have ever been found and, if psychobiology is right, there can be none.

Our initial assumption is of course at odds with the idealistic tradition. We adopt the naturalistic (materialist) view that Platonic ideas, ideas in themselves (Bolzano, 1837), objective knowledge without a knowing subject (Popper, 1972), and objective mathematical reality (Putnam, 1975) are so many figments of the metaphysician's brain (Bunge, 1981a). Just as there is no motion apart from moving things, so there are no ideas in themselves but, instead, ideating brains. To be sure we may *feign* that there are ideas in themselves and in fact we must often make such pretense. We do so whenever we abstract from the real people who think up such ideas as well as from the personal and social circumstances under which they ideate. We may do so as long as we do not forget that we indulge in fiction, whether useful or idle. If in doubt that this is indeed a fiction, imagine the fate of human ideas upon a total nuclear holocaust.

All mental processes are cognitive operations but the converse is false. Thus reading and writing, drawing and often also calculating, involve sensory-motor activities as well as the manipulation of artifacts, in particular learning tools. Over the past four hundred years man has invented and manufactured thousands of learning tools, from microscopes to thermometers, from particle accelerators to high speed computers. All these artifacts are used as adjuncts to the nervous system of somebody, expert or lay: they enjoy no autonomy. Even an automated measuring

device presupposes a design involving hypothesizing, and its supplies food for further thought; and even the most automatic of computations is but a link in a cognitive process occurring in some brain. Just as there is no robot work but only robot-aided work, so there is no automatic cognition but only computer-aided cognition.

Since cognition is a function of the nervous system, we must turn to neuroscience and psychobiology (physiological, developmental and evolutionary psychology) for its explanation. (Pure psychology, traditional or modern, can only supply some data and pose some problems about cognition. Genuine scientific explanation, unlike mere description and prediction, involves some mechanism or other—in this case some neural mechanism.) Now, the psychobiological study of learning suggests that learning consists in a change in the connectivity of some neuronal system or other. To put it negatively: If a behavior is unchanging (stereotyped, genetically programmed), then it is controlled by some system composed of neurons the interconnections of which do not change quickly over time. (To be sure these rigid systems are not unchangeable: they develop from birth on and deteriorate with old age. Moreover, if injured or even destroyed they can often reassemble somewhere else. But the point is that these changes are slow compared to the extremely fast changes in neuronal connectivity that we identify with mental processes.) In short, only *plastic* (or modifiable, or self-organizable, or uncommitted) neuronal systems are capable of learning anything. We call any such system a *psychon* (Vol. 4, Ch. 4).

Our first postulate, then, is that learning is an activity of plastic neuronal systems, or psychons. Our second assumption is that some animals do have psychons. (Exactly which species have them is for comparative psychobiology, or evolutionary psychology, to find out.) Our third, that all the psychons of an animal are coupled to one another forming a supersystem. Our fourth, that every animal endowed with psychons is capable of acquiring new biofunctions in the course of its life. (Apparently in man the processes of dendritic sprouting and formation of new synaptic connections stop only at the onset of senility. So, barring senility—which only about one out of five of us experiences—we can continue to learn new items throughout life.) Finally we decide to call *learned* any neural function involving a psychon that has acquired a regular connectivity, i.e. one that is constant or varies regularly rather than at random. A new connectivity may be formed by chance for the first time. If it consolidates, i.e. if there is recurrence or recall, we regard it as established or learned.

Imagine a group of neurons in a young child's cortex. There are three possibilities: (*a*) the neurons are committed, i.e. they form a (rigid) system or part of one, that is genetically programmed; (*b*) the neurons are available: they can associate in one or more ways, ephemerally or permanently, either in a constant manner or in a regularly varying one; (*c*) such association has already occurred. In the first case we say that the group of neurons in question has a rigid connectivity and therefore *cannot learn* anything; in the second, that it has a potential plastic connectivity and therefore *can learn* something; in the third, that it has already *learned* something, though it can presumably learn further items or unlearn some or all of them.

New neuronal systems can be formed during the lifetime of any animal endowed with plastic neurons; and some such systems may be consolidated by use whereas others can be weakened by disuse: see Figure 1.1. This is the *use–disuse* hypothesis, proposed by Tanzi, adopted by Ramón y Cajal, and refined and exploited by Hebb in his classical work *The Organization of Behavior* (1949). In another seminal work, *A Theory of Intelligent Behavior* (1976), Bindra proposed and applied a substantial modification of Hebb's idea of a neuron assembly: the psychons (which he called *pexgos*) need not have a fixed spatial location but may be itinerant, so that every mental act, even a repetitive one, may consist in the formation of a fresh psychon.

There is considerable empirical evidence for the hypothesis that neurons assemble into multineuronal systems, which in turn are organized into supersystems (Mountcastle, 1978; Szentágothai, 1978). There is also a fair amount of evidence for the hypothesis of plasticity (cf. Flohr and Precht (Eds.), 1981) and, in particular, for the hypothesis that "use" strengthens synaptic transmission. (See, e.g. Bliss (1979) and Goddard (1980) for both short term and long term changes in connectivity.) Finally, there is mounting evidence for the hypothesis that mental processes are processes in plastic neural systems (e.g. Baranyi and Fehér, 1981). All such

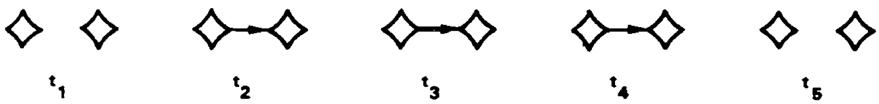


Fig. 1.1. Use–disuse. At time t_1 the two neurons are disconnected from one another. At t_2 one of them synapses onto the other: a system has been formed that may do things neither of the isolated neurons can. By t_3 the connection has been strengthened by use. At t_4 the coupling has weakened through disuse. At t_5 the connection has ceased through prolonged disuse or degeneration (e.g. pruning of dendrites).

hypotheses and data suggest a theoretical framework possessing considerable explanatory and heuristic power (Bunge, 1980a). However, this is just a beginning: we need plenty more detailed neural models, at least one for every cognitive function, and precise tests must be devised and conducted to check them. In any event, if we wish to understand the mind, and in particular cognition, we must give up the traditional arrogant armchair attitude and study the body, for minding is a brain process.

(Several mathematical models of learning psychons incorporate Hebb's use and disuse hypothesis. For example, Amari and Takeuchi (1978) postulate that the learning or self-organizing process consists in a change of the synaptic weights w_i associated to the input signals x_i according to the rate law

$$\tau\dot{w}_i = -w_i + cx_i H(w \cdot x - h),$$

where τ is a time constant, c a positive number measuring the learning efficiency, H the unit step function, h the threshold value, and $w \cdot x$ the scalar product of the synaptic weight vector w by the input vector x. If any of the stimuli are inhibitory, the corresponding synaptic weights are negative. By solving the above system of equations one finds that $w(t)$ is the weighted time average of $H(u)x(u)$ from $u = -\infty$ to $u = t$ with the weighting factor $(1/\tau)\exp[-(t-u)/\tau]$. By assuming that x is a random sequence of a certain type, a number of interesting consequences follow. But, because the neurons in this model are not supposed to fire unless they are excited externally, they are capable of learning only what they are taught. Plasticity, then, is not enough to account for spontaneous learning: if we wish to represent the creativity characteristic of conceptual learning we must include spontaneous neural activity. One way of doing so is to replace the above first order differential equation with a 2nd order one.)

In sum, we assume that the nervous systems of animals of certain species have plastic subsystems, i.e. neural systems with modifiable connectivity, as well as rigid neural systems. We also assume that every mental function is a process in some plastic neural system. For example the formation of a polysyllabic word, or of a sentence, of the form *abc*, probably consists in the sequential activation of neural units A, B, C, in the same order. This speculative hypothesis explains also thought and speech anomalies such as spoonerisms. Thus saying 'pointcounter' when intending to say 'counterpoint' might be due to the wrong timing of the psychons engaged in the production of 'counter' and 'point'. And in turn this time reversal might be

caused by the temporary inhibition, or else the delay, of the first psychon, which inhibition (or delay) could be due to a number of factors, such as a strong coupling between the delayed psychon and a third psychon, or to external interference.

Our hypothesis that the nervous system is the organ of cognition is of course the very first postulate of any biologically oriented psychology and epistemology. This view is at variance with the fashionable cognitivist school and its computational or "functionalist" philosophy of mind (Putnam, 1960; Fodor, 1975, 1981; Pylyshyn, 1980). These are the basic tenets of cognitivism:

(i) Cognition is computation, and as such it is not a prerogative of brains: it can also be performed by computers. (Yet many cognitive processes, in particular perception and imagination, are not "intentional rule-governed phenomena" but occur spontaneously and unconsciously.)

(ii) Computation takes place in the mind, which is a kind of computer. (By pushing computation into an immaterial mind, the study of it does not fall under the jurisdiction of natural science. The fact that some computations can be simulated on computers does not prove that computation *in vivo* is essentially the same as its artificial mimicry.)

(iii) Any suitable theory of cognition is basically a theory of Turing machines. (This theory is so extremely simple that it cannot account for the humblest thing in the universe, such as an electron or a photon—let alone for a brain with its myriads of neurons and their plastic connections.)

(iv) Knowledge of the brain is irrelevant to understanding cognition and, in general, matter does not matter to mind. (Since cognitive processes are brain processes, neuroscience is essential to the advancement of cognitive psychology.)

(v) Artificial intelligence is the proper foundation for cognitive psychology, not the other way round. (So far this approach has produced only analogies. It is to be expected that, as neuropsychology advances, it will be able to guide work in artificial intelligence for, after all, imitation presupposes knowledge of the real thing.)

For the reasons advanced in parentheses, we regard cognitivist psychology not only as mistaken but as wrong-headed and paralyzing: as a collection of simplistic and superficial metaphors, a modern version of psychoneural dualism resulting in a further estrangement of psychology and epistemology from science, in particular biology. (For further criticisms see Bunge (1980a) and Vol. 6.)

1.2. *Brain States and Processes*

We have assumed that *every cognitive process is a process in some plastic subsystem of the nervous system of an animal.* Consequently an animal lacking a nervous system—be it small like an amoeba or large like a sponge—cannot know anything. It can of course react to stimuli of certain kinds and even act spontaneously—but so can any lump of matter.

We also assume that the nervous system, far from being a mushy thing, is a highly organized biosystem composed of subsystems each of which discharges certain peculiar functions in addition to the functions common to all living systems (such as metabolism and protein synthesis). Sometimes the specialization goes down to the individual neuron level. Thus the neurons in the mammalian visual cortex are so specialized that, whereas some of them detect only vertical lines, others react only to lines tilted at a 20° angle, and so on; still others detect short lines, wide lines, or what have you. (See Hubel and Wiesel, 1977; Gilbert and Wiesel, 1979.) In short the individual neurons of that kind analyze the incoming sensory inputs simply by selectively detecting some of their features. The integration or synthesis of these specialized activities is done by multineuronal systems. Traditional neuropsychology held that the brain is a sensory analyzer: we know now that it performs both the analysis and the synthesis of incoming stimuli, and also that it acts as a shock absorber for most of them, particularly the repetitive signals.

Whereas some cognitive processes occur in rigid (committed, wired-in, programmed) neural systems, others engage plastic neural systems as well, or perhaps exclusively. Thus the mere detection of ordinary stimuli requires only rigid neural systems, and so does the classing of stimuli into harmful, harmless, and beneficial. On the other hand locations or times, faces or letters, tricks or skills of any kind, presumably take plastic neural systems or psychons. The latter have been experimentally identified in a few cases. Thus Lindauer (1971) trained bees to forage at certain places, and then transplanted their tiny mushroom bodies into the brains of untrained bees. After recovery the latter made a bee line for the feeders. Knowledge can then be surgically removed and transplanted into a different animal—much to the chagrin of the mentalist.

The hypothesis that every cognitive activity is a process in a specialized neural system (i.e. the specific activity of it) can be elucidated as follows. To say that thing x is *active* at time t is the same as saying that at least one property of x is changing at t. Now, every property is lawfully related to at

least one other property of the same thing: i.e. there are no stray properties. (For this ontological principle see Vol. 3, Ch. 2, Section 3.3.) So, if one property of a thing changes, other properties of the same thing will change as well: i.e. every change is multidimensional. For this reason we should always consider clusters of properties and, in principle, all the properties of a thing. If we do so we can speak of the states and the changes of state of the thing of interest. Every *state* is a list of properties, and every change of state of a thing is called an *event* occurring in the thing. The sequence or list of all the changes undergone by the thing during a time interval is called the *process* occurring in the thing over that interval. (There are no events or processes in themselves: every change is a change in some thing. Nonthings, such as immaterial souls and disembodied ideas, are in no state and *a fortiori* cannot change state.)

(A general and useful way of exactifying these ideas is as follows. First, represent each property of a thing of kind K as a mathematical function, in principle a time dependent one. (Never mind the other dependent variables for the moment.) Call F_K^i the function representing the ith property of thing of kind K. Next, collect all these functions into a list or n-tuple $F_K = \langle F_K^1, F_K^2, \ldots, F_K^n \rangle$, called the *state function* of an arbitrary member of K. The value $F_K(t)$ of the state function at time t may be pictured as an arrow representing the *state* of the thing at time t. As time goes by the tip of the

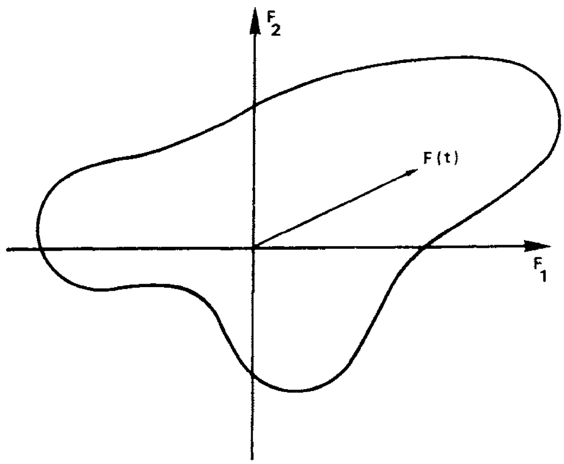


Fig. 1.2. State space for an (imaginary) thing characterized by two properties, represented by functions F_1 and F_2, combining into the state function $F = \langle F_1, F_2 \rangle$. The tip $F(t)$ of the "vector" F at time t represents the state of a thing of the given kind at t.

"vector" $F_K(t)$ spans a region of an n-dimensional space called the *state space* for things of kind K. (See Figure 1.2 and, for details, Bunge (1977a) or Bunge (1979a).) In other words, the "life history" of a thing of kind K is representable as a trajectory (or a bundle of trajectories) in the state space $S_K = \{F_K(t) | t \in T\}$, where T, a subset of the real line, is the set of all instants. The laws of things of kind K restrict the possible values of the state function, and therefore limit the state space for K's to a subset of the set of all logically possible states.)

(A state space for a species K represents all the lawful states of unspecified members of K. To individualize any state of a particular member of K we must specify its particular constraints and circumstances. Call F_K^a the state function representing an individual a of kind K. Knowing all the values of a particular F_K^a over a certain time interval amounts to knowing the history of the individual concerned during that interval. More precisely, we may say that the *history* of an individual of kind K over the time interval τ is representable as the graph of its F_K^a during τ. See Figure 1.3.)

(Now, the history of a thing may be eventful or nearly uneventful during a given interval. I.e., the state function F_K^a for a given individual a of kind K may change considerably or only negligibly over a given period. If the history is eventful we call it the *activity* (or process) of the thing concerned. Speaking formally, the activity of, or process in thing a during period τ, is

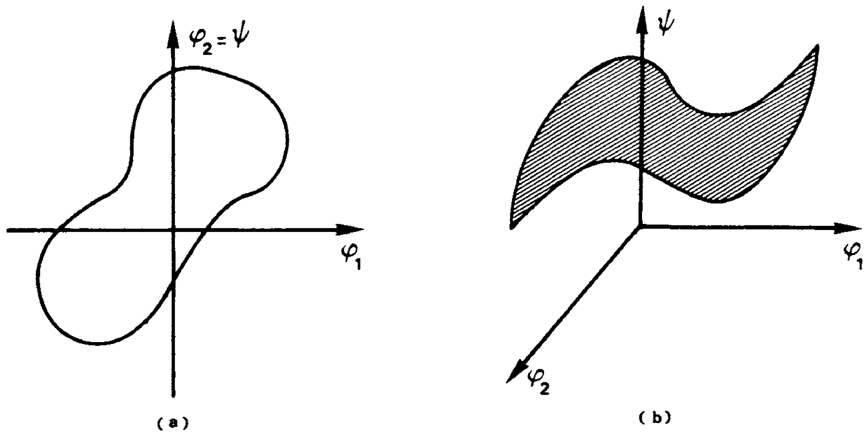


Fig. 1.3. Projection of state space of a plastic neural system. In (*a*) one of the biological properties (axes) is identical with a psychological one. In (*b*) the psychological property is some function of two biological properties—but is itself a biological property.

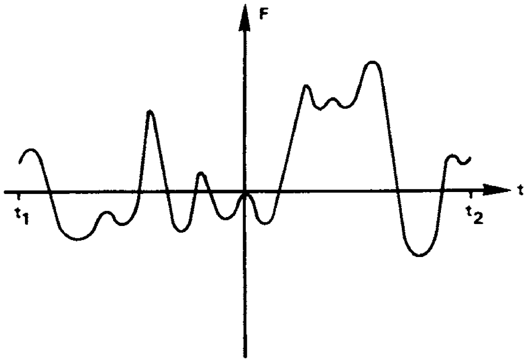


Fig. 1.4. History of a thing during the time interval $[t_1, t_2]$.

representable by the set of ordered pairs $\langle t, F^a_K(t)\rangle$ for t in τ, such that the rate of change $\dot{F}^a_K$ differs from zero:

$$\pi(a,\tau) = \{\langle t, F^a_K(t)\rangle \mid t \in \tau \ \& \ \dot{F}^a_K(t) \neq 0\}.$$

This graph represents the *total* activity of a over τ. See Figure 1.4.)

(In order to represent only the peculiar or specific activity of a thing we must subtract whatever other things do over the same period. Since we are interested in the specific activities of certain subsystems of an entire animal, we shall take "the other things" to be the parts of the animal other than the system of interest. To indicate that x is a *part* of b we write '$x \sqsubset b$' (Vol. 3, Ch. 1, Section 1). Thus the *specific* or *peculiar* process (or *activity* or *history*) going on in subsystem a of animal b during τ is representable by

$$\pi_s(a,\tau) = \pi(a,\tau) - \bigcup_{x \sqsubset b} \pi(x,\tau), \quad \text{with} \quad x \neq a,$$

where, for any sets A and B, $A - B = A \cap \bar{B}$, '$\cap$' stands for intersection and '$\bar{B}$' for the complement of B. In other words, the specific function (or process) of system (e.g. organ) a in animal b equals its total function minus the total functions of all the remaining parts of the same animal. Thus imagine a system with only two components, and call them a and b. Suppose that their respective activities or processes during a given time interval are $\pi(a) = \pi \cup \pi_1$ and $\pi(b) = \pi \cup \pi_2$. Obviously their specific activities are $\pi_s(a) = \pi(a) - \pi(b) = \pi_1$, and $\pi_s(b) = \pi(b) - \pi(a) = \pi_2$.)

For example, seeing is the specific activity of the visual system, which in the higher vertebrates includes the visual cortex and is coupled to the motor system. (Every sensory system is a subsystem of its corresponding sensory-

motor system. Thus looking is accompanied by saccadic movements, sniffing by puckering the nose, tasting by moving the tongue, listening by turning the head, and touching by moving a limb. After all, the sensory stripe is right next to the motor stripe in the cerebral cortex. In short, there is no perceptual learning without motor activity. In other words perceptual learning is not just a matter of receiving passively sensory inputs, but of combining a selection of such inputs with motor outputs: see Held and Hein (1963).) Likewise making a computation is presumably the specific activity of some psychon(s) in the association cortex; becoming conscious of one's own thoughts or feelings is presumably the specific activity of a neural system monitoring other neural components of the same brain, and so on.

Neuroscientists and psychobiologists are working on the problem of mapping the brain, in particular the cerebral cortex—i.e. of finding out what (neural system) does what (function or activity). This task was started by Penfield (1950) who, by stimulating electrically certain points of the cerebral cortex of waking patients, evoked in them various perceptions, images, memories, and thoughts. A suitable map of the brain would equate every cognitive function with a particular activity of some neural system or other (cf. Olds, 1975). Note that we have written *equate* not just *correlate* each cognitive function with a neural activity. Whereas the former expression suggests the identity of cognitive function and specific brain function, the latter expression suggests only a vague association between cognition and brain functioning in the style of psychophysical parallelism or interactionism. The dualist, to whom the mind is not a collection of brain functions but a separate substance, has no real use for brain mapping.

This mapping is such a formidable enterprise that it may be doubted whether it will ever be completed (Hebb, 1980). Consider: the plastic part of an individual human brain is composed of at least 1,000 million neurons. If each psychon has between 100 and 10,000 neurons, the total number of psychons is at least one million—and many times this figure if psychons are itinerant rather than fixed. Moreover it is well known that there are considerable differences among individual brains, so that the precise location of a particular cognitive function in one person is at most a very rough indication for its location in the next person. Nevertheless there has been considerable progress in this area over the past few years, as new techniques of imaging brain processes have been developed. These techniques have laid to rest the holistic hypothesis that the brain functions either as a totality or not at all.

TABLE 1.1
Cognitive deficits or disorders and their plausible neurophysiological explanation

Name or description	*Plausible explanation*
Blindness	Visual system damage or malfunction
Impaired vision	Multiple sclerosis, brain tumor, or dioptric defect
Visual agnosia	Lesion in occipital or posterior temporal regions or left hemisphere
Visual illusions and hallucinations	Lesions at any level in the visual system from eye to cortex
Deafness	Auditory system damage or malfunction
Unawareness of left half of body	Right parietal lobe damage or malfunction
Loss of attention	Reticular activation system or thalamus damage or malfunction
Amnesia	Bilateral hippocampal or diencephalic damage or other lesions or diseases
Loss of short term memory	Hippocampus damage or malfunction
Loss of curiosity	Frontal lobe damage
Loss of consciousness	Reticular activating system damage or malfunction
Aphasia	Wernicke or Broca area damage or malfunction
Dyslexia	Gyrus lesion
Agraphia	Temporal–parietal occipital lobe damage, retrorolandic lesion
Acalculia	Temporal or parietal lobe damage

Table 1.1 lists some of the cognitive deficits or disorders most frequently "seen" by clinical psychologists, psychiatrists, neurologists, and neurosurgeons. (See Pincus and Tucker, 1978; Hécaen and Albert, 1978.) We seem to know more about brain abnormality than about normality for two reasons: (*a*) the study of malfunction is one of the main ways of finding out what the normal function of a component is: we notice the light bulb in the dark staircase when it has burned out; (*b*) there are more researchers and means engaged in medical research than in neurobiological research—because of a science policy inspired in a pragmatist philosophy of science and technology.

Many workers speak of *information processing* where we speak of *brain processes*. That mode of speech is justified insofar as the nervous system does possess components capable of detecting, absorbing, transducing, or propagating signals throughout the system. However, it does not follow that the statistical theory of information is actually used to advantage to describe or explain such processes, and so no specific neural models need be constructed and tested. In fact information theory (Shannon and Weaver, 1949) applies only to messages transmitted through noisy channels and having definite probabilities of being received and decoded, whereas neuroscience and psychology are interested in *specific* biophysical, biochemical and biological processes occurring in the nervous system. When neuroscientists say that '*A* sends information to *B*', or that 'There is a flow of information from *A* to *B*', they ought to say simply that *A acts on B through C*. And when they write '*B* processes information coming from *A*' they ought to say simply: *A acts upon B*, triggering in *B* a process peculiar to *B* and different from any processes in *A*. After all they are dealing with systems of a very special kind, namely nervous systems, not—like the information theorist—with all kinds of information processors.

Besides, the brain is not just a processor of incoming stimuli: it also generates signals. True, S-R (or behaviorist) psychology, as well as "ecological" psychology (Gibson, 1979), hold that the brain does not create anything but picks up information from its environment. But of course any neuroscientist knows that the brain of the higher vertebrate has a constant autonomous activity. And any psychologist might argue that the view that there is no creation of new information leads to infinite regress. Indeed, if all I know is what I learned from *A*, who in turn learned it from *B*, and so on, then the only way of putting an end to the chain is to postulate either a Prime Mover or a Platonic realm of eternal ideas, neither of which is a useful and testable hypothesis. Besides, we should recall that, although

the environment of mankind has always contained electromagnetic fields, molecules, and cells, it took millions of years of biological and cultural evolution for the emergence of organisms capable of creating the corresponding ideas. In short, information is not out there for grabs: the organism absorbs most of the stimuli impinging upon it, transforms a few of them into processes of other kinds (i.e. transduces them), and generates its own, part of which it transmits to other animals.

2. Cognitive functions

2.1. *Perceiving, Thinking, and All That*

We get to know about the world and ourselves through perception and ideation, two distinct but intimately related neural processes. Perception is the last stage of a process that normally starts as mere detection or sensation in some neurosensor (or neurodetector), such as a nerve ending in the skin or in the tip of the tongue. Sensation is the specific activity of sensory systems such as the olfactory and the auditory systems. The neuroreceptors are quite unlike physical or chemical receptors, in that they are under the constant "downstream" action of the central nervous system. Consequently the state of a sensory system depends on the state and history of the brain as well as on the nature and intensity of the stimulation. Hence we are unlikely ever to sense the same stimulus twice in exactly the same manner. Yet we may perceive the two stimuli as identical even if they are not. Moreover we may miss a tone or a motion altogether unless we were expecting it. (See Figure 1.5.) To put it into other words: the response is a function of both stimulus and internal state of the organism. (When Hunter said this in 1929 he was not listened to: the behaviorists were prepared to hear only results fitting the stimulus-response model, which ignores the state of the organism.)

In the higher vertebrates sensation is processed in the sensory cortical "area", which is divided into the primary, secondary, and tertiary "areas". The primary "area" is a component of the corresponding sensory system, and apparently it is particularly malleable during the early stages of development, which is when the animal learns to smell, see, hear, etc. We shall assume (Hebb, 1966; Bindra, 1976) that, whereas sensation is the specific activity of a sensory system including the primary sensory cortical "area", perception is this activity together with the specific activity of the plastic neural systems directly attached to it, i.e. the secondary and tertiary

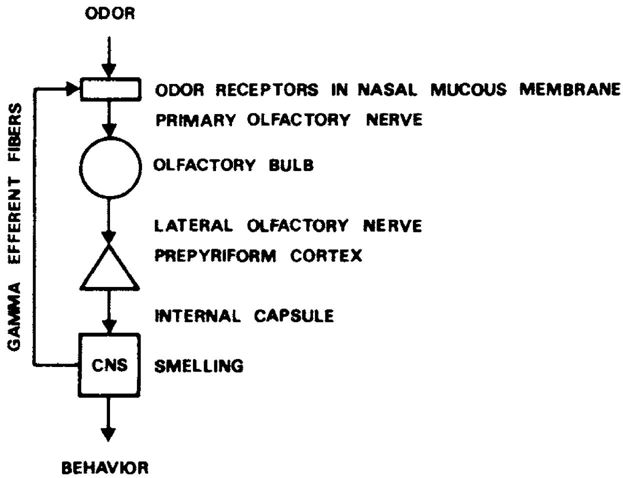


Fig. 1.5. Smelling, a primary cognitive activity, as a nervous system process. The diagram sketches the large scale organization of the olfactory system in mammals. Note the distance between odor reception and odor perception (smelling). Note also the feedback of the CNS (central nervous system) on the receptors.

"areas", as well as whatever other plastic neural systems may be activated by them. Thus there is a long neural way between odor (or light or sound or pressure) reception and its perception, i.e. smelling (or seeing, hearing, or feeling).

Far from being isolated from other neural activities, perception is linked with a number of them, including ideation and behavior. (See Figure 1.6.) This link accounts for the fact that perception can occur in the presence of fragmented stimuli, on the basis of minimal cues, and even in the absence of sensory stimulation. It also accounts for the various constancies (of, e.g. distances and shapes) as well as for spontaneous perceptual processes such as hallucinations and those involved in dreams. This view is at variance with the causal theory of perception, according to which every percept is caused by an external thing, and this in an invariable manner. This view, which is still going strong, ignores the ongoing spontaneous activity of the brain as well as the complexity of the processing of signals coming from the sensors. Perceiving is never a passive effect of the thing perceived: it is not copying but constructing (Neisser, 1967; Bindra, 1976), and therefore it is midway between sensing and thinking (Gregory, 1970). When perceiving something we construct a percept of it with sensations, memories, and expectations.

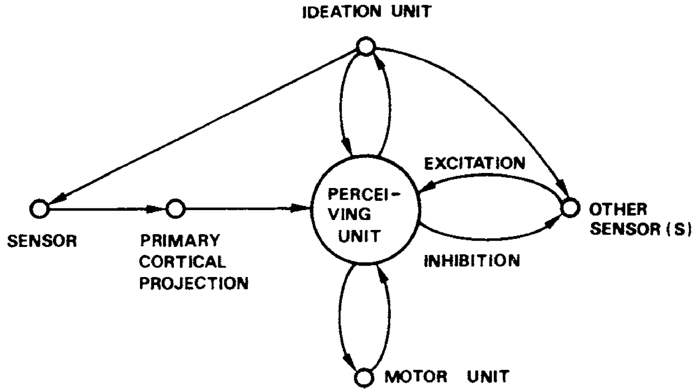


Fig. 1.6. Perception is linked to sensation, ideation, and behavior. Schema inspired by Bindra (1976), who in turn modified it from Hebb (1949).

In other words, the stimuli coming from the perceived object elicit *changes in the ongoing activity* of the nervous system rather than causing the latter. (This is why we assume that an animal perceives an external object as the symmetric difference between the activity of the corresponding process in the presence of the object, and in the absence of it: see Bunge, 1980a, Postulate 4.1.) Therefore there can be perception without external stimuli, as in hallucinating and dreaming, but not without a perceptual system. It follows that there can be no extrasensory perception (ESP). And if there is no ESP (clairvoyance, telepathy, precognition, and the like), then parapsychology is an illusion (Wundt, 1879; Hansel, 1980). What remains to be explained is not the alleged facts of parapsychology but the gullibility of parapsychologists—a task for social psychologists (Alcock, 1981).

Strictly speaking we do not perceive things but events (changes) occurring in things. Thus we cannot see an object unless it emits or reflects photons of a certain energy. Moreover some such events, to be perceptible, must trigger processes ending up in some neurosensor or other. Unless the two conditions are jointly satisfied the object remains invisible. And the way the animal perceives the events occurring in its neurosensors is by *mapping* them into events in some of its perceptual systems. Admittedly this mapping is partial (not everything gets represented) as well as distorted. Nevertheless it is a genuine map, i.e. a representation of certain sets (of events) into another set (of events). The first set (or domain of the function) is a collection of events in the environment or the periphery of the animal, the second set (or codomain) is a collection of events in the perceptual

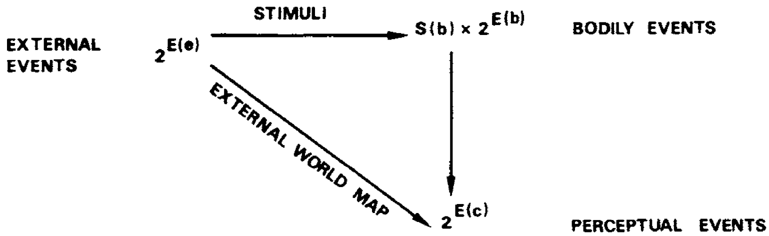


Fig. 1.7. Events in the environment e of animal b are projected into pairs ⟨bodily state, set of bodily events⟩. This map dovetails with the next, or somatosensory map, which projects external events into cortical events, or points in the secondary or tertiary sensory area. The composition of the two maps yields the external world map for the corresponding sensory system. $E(e)$ is the set of possible events in the environment of the animal; $S(b)$ the state space of the animal and $E(b) \subset S(b) \times S(b)$ its events space. $2^{E(b)}$ is the family of all subsets of $E(b)$; $2^{E(c)}$ is the family of all subsets of the event space $E(c)$ of the cortical system concerned. Details in Bunge (1980a).

system of the animal. In every case the map or function depends not only upon the sensed stimuli but also upon the state of the animal. See Figure 1.7. Maps or internal representations of this kind allow the animal to "navigate" in its environment.

Whereas some memories are remembered, others are forgotten, temporarily or permanently. And those that are recalled are so, more often than not, in a distorted way: like perceiving, memorizing is creative—i.e. not just storing and retrieving but reconstructing (Bartlett, 1932). Donald Hebb used to ask his students to remember breakfast. They all "saw" themselves (with their mind's eye, as we say in ordinary language) sitting at table—an obvious construction. This constructive or creative feature of memory is ignored by the computer (or information-theoretic) metaphor, according to which the learned items are encoded, stored, and then retrieved, possibly with loss, never with gain. This holds literally for computer memories and library systems, but is grossly off the mark in the case of brains.

Since the storage models of memory are inadequate, some dynamical model is likely to prevail. One such model involves the following hypothesis: "Remembering consists in the fresh production or reconstruction of an item, not in the *re*-activation of some permanent representation of the item as such, and what item is remembered—the response production—is the pexgo [psychon] selectively activated at the time of remembering" (Bindra, 1976, p. 330). Rather than permanent changes in

circuitry, there may be a propensity or disposition for certain neural systems to be activated, either spontaneously or induced by others. Such dispositions may be quantitated as the probabilities of certain synaptic connections. Any such probability increases with use and decreases with disuse or abuse (e.g. drugs), or just age. The ability to memorize is of course necessary for learning anything: patients who have lost their short term memory (as a result e.g. of brain injury) are incapable of learning anything but for some elementary behavior patterns. (Amnesic patients can be conditioned even though they fail to remember the conditioning process: Weiskrantz and Warrington (1979).)

Of all the brain functions thought is the most remarkable and the least understood. Forming concepts, transforming and interrelating them to construct propositions, questions and directions, are examples of thinking. (Many psychologists include visual images among thoughts—but then why not auditory, olfactory, and tactual images? We regard visual imagination as perception in the absence of sensory stimulation, and submit that it can help or hinder thinking, not however take its place.) Let us deal quickly with the formation of concepts and propositions.

One of the characteristics of the formation of concepts, as well as of the production of the corresponding verbal expressions, is that they may occur in the absence of external stimulation: they can be purely internal processes without sensory inputs or behavioral (motor) outputs. And, when elicited by external stimuli, concept formation (and the corresponding speech production) may involve lack of discrimination. Thus we can form the general concept of a tree perhaps because we possess plastic neural systems the specific function or activity of which is roughly the same whether they are activated by seeing or remembering a maple tree, a fir tree, or a birch tree: such psychons are generalists rather than specialists (Bunge, 1980a). An alternative account of concept formation is this. Every time an animal perceives a tree, several million neurons are excited, perhaps a different set every time. But there may be a thousand neurons common to all those different excitations. These will become organized into a neuron assembly, for neurons that fire together tend to stay together. The activity of this neuron assembly is (identical with) the general idea of a tree (Hebb, 1949). It is up to experiment to determine whether one of these hypotheses, or a third one, is true.

We hypothesize that forming complex concepts, propositions, problems, instructions, and other composite ideas consists either in psychon grouping or in sequential activation of two or more psychons (possibly cortical

columns). Thus the concept of an even number would be formed by the pairing of psychons for "number" and for "even". An alternative hypothesis is that such a concept consists in the successive activation of the simple psychons concerned—an extremely fast succession indeed. (In this case language would certainly make a difference for, whereas in English and German 'even' comes before 'number', in the Romance languages it is the other way round.) However, the two hypotheses are mutually compatible. Indeed, psychon pairing may occur during learning (or relearning), whereas sequential activation may happen while using or applying the learned item, since language may be regarded as a vast storehouse of ready-made thinking bricks. The young child has to think hard to add two and three: we just use the result, encoded in a linguistic entity, when needed. (More on language and thought in Ch. 3, Section 3.2.) Once the learning period is over we can perform uncounted mental operations in a thoughtless ("mechanical") manner, just as we walk or drive relying mostly on automatisms. But in either case we may have to resort to new thinking when faced with novelty. (Unfortunately we do not always do so: witness the ideologue who, because he cannot cope with novelty, wants the world to stand still.)

Psychon pairing and sequential activation explain much but not everything. In particular they do not account for our occasional ability to direct thought or "concentrate" it on one task in spite of external (sensory) and internal (spontaneous) excitations that could be diverting. However, the directedness of thought is explained psychologically by the fact that it always occurs in a context: that of the problem at hand (or rather brain), which in turn is embedded in a body of knowledge. In turn, this context has a possible physiological explanation: it could be the background activity of psychons associated with those doing the thinking (Hebb, 1976). And the fact that we are able to restrict the activity to some psychons (rather than engaging the entire plastic brain) can be explained in terms of lateral inhibition, or the containment of the spreading of nervous excitation. To be sure all this is speculation, but reasonable and fertile: hypotheses such as these underlie empirical research into the physiology of thinking. (See Bechtereva, 1978; Goddard, 1980.)

In any event we assume that a thought, be it a concept, proposition, inference, problem, or command, is a brain process. More particularly, we postulate that every thought is a specific process occurring in some plastic neural system or psychon of some animal. Hence a thought may be described as the sequence of states of such psychon over the time lapse it

takes to be completed. The particularities of such process are likely to vary from individual to individual, and even from moment to moment in one and the same individual. Thus it may well be that nobody undergoes exactly the same brain processes when thinking of the number 5 at different times: it could be that different psychons are activated at different times (Bindra, 1976). However, any such thoughts of the same construct must be equivalent in an important respect, i.e. the processes must have the same general shape, for otherwise they would not consist in thinking of the number 5.

(Suppose neuropsychology discovers the exact nature of such equivalence, and call it $\sim$. Further, call $\theta_{ij} = \pi_s(b_i, \tau_j)$ the specific process occurring in brain b_i, during period τ_j, when thinking of a certain construct. Since all the processes consisting of thinking of the same construct are $\sim$-equivalent, we write: $\theta_{ij} \sim \theta_{mn}$ for all i, j, m, n. We now form the equivalence class (set of all $\sim$-equivalent processes) $[\theta_{ij}]_\sim$. Finally, we identify the construct c with this equivalence class, i.e. we set

$$c = [\theta_{ij}]_\sim .$$

In other words, *a construct is not an individual brain process but an equivalence class of brain processes* occurring in different brains or in the same brain at different times (Bunge, 1981a). As long as the exact nature of the equivalence relation $\sim$ is not defined, this is a programmatic or heuristic hypothesis that can guide the very research into the identification of $\sim$. More on such equivalence in Ch. 2, Section 1.1.)

The psychobiological conception of thinking explains the association of ideas, that mainstay of empiricist psychology and epistemology, as well as its *bête noire*, intellectual creativity. Association is explained as the interconnection of psychons, and originality as the formation of radically new psychons. Neither is mysterious: interconnection because the whole brain is a system, so that anything happening in one of its components influences some other components. (The influence may be positive and may get established, as in the case of association, or it may be negative or inhibiting, as when every time we think of mathematics we tend not to think of sports.) And originality is not mysterious because of the remarkable spontaneity and plasticity of the cortical neurons.

(The ideas of association and of originality may be formalized roughly as follows. Let S_i be the collection of cognitive acts that subject i performs up until a certain time: call it the *cognitive space* of i. Further, call x and y any two members of S_i, and interpret their concatenation $x \circ y$ as "cognitive act

x associates with (or is followed by) cognitive act y". Finally, call 0 the act that does not increase knowledge even if it was intended to do so; more precisely, 0 represents an arbitrary unsuccessful cognitive act—not to be mistaken for $\varnothing$, or total ignorance. We postulate that the structure $\langle S_i, \circ, 0 \rangle$ is a noncommutative monoid. This amounts to the hypothesis that there are no isolated cognitive acts, but any cognitive act may be associated with some other cognitive act. And we may stipulate that a cognitive act is *original* with subject i if, and only if, it belongs in S_i but not in the cognitive space of any other animal up to the given time.)

2.2. *Learning*

Learning can be conceived broadly or narrowly: as adaptation or as neural modification, temporary or permanent. In the former case one may say that an organism which has developed immunity to some antigen has "learned" to face the latter, and even that a population which has become adapted to a changed environmental—after a wasteful selection process—has "learned" to live in it. Although such a broad construal of learning puts it in a biological perspective, it erases the specificity of the individual neural process accompanying the acceptance or generation of new information. (The same charge can be laid against the information processing view.) Therefore we adopt the narrow construal of learning as neural modification.

We assume, then, that an animal *learns* something if it forms at least one new psychon (plastic neural system) or combines in a novel fashion two or more psychons that had been established previously. And we call *knowledge* of an animal at a given time *the set of all the items it has learned* up until that time—i.e. the collection of changes in its plastic neural supersystem. (Such changes need not all consist in permanent alterations of connectivity: some may consist in changes in the probabilities of certain synaptic connections.) Needless to say, accumulating experience (learning) is not passive recording but a process where signals are selected (mostly filtered out) and transformed—distorted and enriched. In short, learning is *creative* even when it consists only in learning what other animals have learned before. It is creative in the sense that it is not automatic and that it is new to the individual.

Like any other brain process, learning can be represented as a change in an abstract space representing the possible (lawful) states of the brain. An animal engaged in a learning process can be said to expand its plastic neural

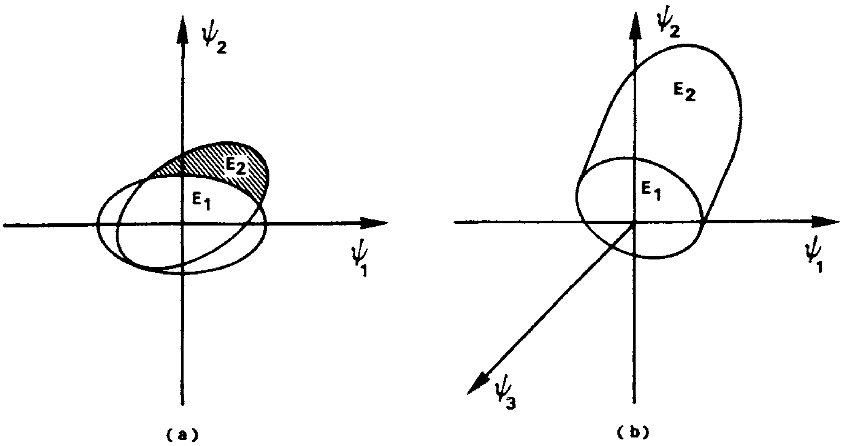


Fig. 1.8. Two kinds of learning. (*a*) Learning of one more item of a given kind (with eventual loss of some experience). (*b*) Learning something qualitatively different is represented by the intervention of a third coordinate axis in the neural state space. Adapted from Bunge (1980a).

state space. This extension can be quantitative or qualitative: the former when the animal learns one more trick of a kind it has sampled before, and the latter when it learns something of a different kind. Thus learning one more word and one more algorithm fall into the first category. On the other hand the learner's neural state space expands qualitatively—i.e. new axes sprout—when it masters a new subject for the first time. The most rewarding aspect of research and advanced teaching is that it encourages the sprouting of new axes in one's brain state space. See Figure 1.8.

Several learning theories are compatible with both the neuroscientific and the psychological data. Only a few of them are deep, i.e. concerned with changes in neural systems, but they are not detailed enough. Most learning theories are purely descriptive: they disregard brain mechanisms and concentrate on the necessary and sufficient environmental conditions for learning. (In particular, the fashionable information-processing view, which ignores the nervous system, is a metaphorical description of what happens during the learning process.) However, some of these theories do capture certain important features of learning, particularly that it is a stochastic process, i.e. a sequence of events every one of which has a definite probability.

Figure 1.9 illustrates the class of stochastic learning theories in the case of single trial learning. The animal is assumed to have probability p of

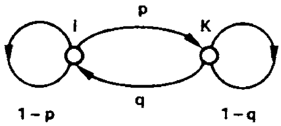


Fig. 1.9. A stochastic model of all-or-none learning events. The arrow from I to K symbolizes learning, and the converse forgetting. The closed loops represent not learning (left) and not forgetting (right).

learning a given item at a trial, hence probability $1 - p$ of not learning it; and it has probability q of forgetting it, *ergo* $1 - q$ of remembering it. If multiple trials are allowed a well known pattern emerges. For any given task the number of successes increases with practice (i.e. number of trials) as well as with the number of successes. (Successful learning is its own reward.) Likewise the number of errors decreases with practice as well as with the number of errors that get corrected. (Otherwise the process is haphazard, i.e. no trend emerges.) The result is the logistic curve: Figure 1.10. This curve summarizes a purely descriptive theory of learning that does not account for the learning mechanism—let alone its neurophysiological "substrate" or "correlate". It is a datum to be explained by neurophysiology.

Another view of learning that fails to make contact with neurophysiology is innatism, according to which "Learning is primarily a matter of filling in detail within a structure that is innate" (Chomsky, 1975, p. 39). On this view knowledge of a certain kind, i.e. pre-existing—in

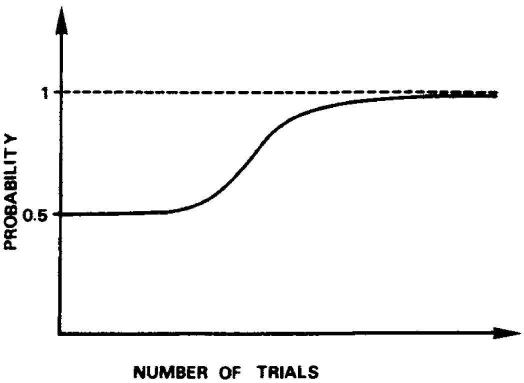


Fig. 1.10. The logistic learning curve: if errors are corrected, the probability of correct performance approaches 1 as the number of trials increases.

particular innate—knowledge, makes learning possible, not the other way round (Chomsky, *op. cit.*, p. 118). Needless to say, there is not the slightest shred of empirical evidence for this speculation. On the contrary, the anatomy, physiology and psychology of development show that the newborn human knows nothing, learns only gradually, and cannot skip any stages because learning keeps pace with the maturation of the nervous system.

The existing learning theories are nonspecific: they are assumed to represent motor learning (e.g. learning to ride a bicycle), perceptual learning (e.g. learning to discern a pattern in an assemblage of dots), conceptual learning (e.g. learning new concepts and propositions), etc. Yet it is obvious that there are big differences among these various kinds of learning, and it is well known that they are done with different subsystems of the brain. Thus conceptual learning, unlike learning of other kinds, involves understanding and it may proceed without any sensory-motor concomitants. So, in addition to general learning theories we need learning models representing each a given type of learning. For example, a model of motor skills learning, even if superficial (i.e. non-neurophysiological), should account for the interplay of behavior and cognition. Thus in the higher vertebrates stimuli that can teach (unlike those that elicit only reactions of the knee jerk type) do not trigger responses automatically but elicit the intervention of cognitive activities such as anticipation and recall.

In other words, the learning process initiated by an environmental stimulus is not summarizable by a stimulus-response arrow. Instead, what happens is that the stimulus is paired to some cognitive state, and both produce jointly a response. (In short, instead of $s \mapsto r$ we have $\langle s, c \rangle \mapsto r$.) In turn, the response has some consequence or outcome (internal or external). If the animal receives information about such outcome, this information may change its cognitive state, producing next time a different response, hence a different outcome, in the presence of the same stimulus. In short,

First trial $\langle$stimulus, cognitive state$\rangle \mapsto$ response $\mapsto$ outcome.

Second trial $\langle$stimulus, new cognitive state$\rangle \mapsto$ new response $\mapsto$ new outcome.

By comparing the successive outcomes the animal may learn whether they converge to the goal or desirable state. If they do not, it may correct the strategy—or give up. This sektch of a feedback learning model can be mathematized, "embodied" in a neural system, and in principle tested both behaviorally and neurophysiologically.

Not only learning theory but most of psychology and the whole of epistemology are attempts to answer the question 'How do we learn?' The empiricist answer is: 'I learn from my own experience and provided my behavior is reinforced (e.g. my ideas are confirmed)'. Skeptics and idealists, on the other hand, deny that we learn from experience—unless perhaps from negative experience, i.e. from mistakes. As is so often the case, each of the contenders holds a grain of truth. We do learn something (not everything) from experience—provided our nervous system is in a receptive condition, we are motivated (and thus willing to make an effort), and the material to be learned is somewhat organized.

That the nervous system must be in a receptive condition is obvious, although this point is ignored by both empiricists and idealists. No matter how well a learning material is presented, a person under the influence of alcohol or some other drug may be unable to learn it. But a healthy brain is not enough, nor is skillful teaching sufficient. Thus whereas Pavlov believed that we can learn automatically from paired experiences (e.g. sound of a bell and temporally contiguous presentation of food), it has been found that repeated paired experiences do not guarantee learning. (See Dickinson and Mackintosh, 1978.) It has been conjectured that, for somebody to learn to do B whenever A happens, he must somehow realize that A and B may be associated (Dawson and Furedy, 1976).

Learning requires also some motivation: an animal without motivation to learn will not explore (Bindra, 1974). Motivation for learning can be *intrinsic* (curiosity) or *extrinsic* (for the sake of a reward other than satisfying curiosity). Hence we can teach ourselves, and others can teach us, to regulate both motivation sources. Thus the standard teaching techniques, by emphasizing extrinsic reward (praise, prize, raise in status, etc.) may inhibit adventurousness and may favor a mercenary attitude towards learning. (Anyone trained to respond primarily to external rewards will say to himself that, since he wants a sticker or some other prize, he should pick problems he knows how to solve safely and quickly rather than problems that will keep him wandering and wondering.) In any event, learning is the easier the more strongly motivated—and motivation is in turn enhanced by learning.

Learning, then, is not just picking up information and processing it in a mechanical way: it is a highly selective and creative process (Piaget's *assimilation*). Nor does the brain remain unaffected by learning: it restructures itself as a result of learning (Piaget's *accommodation*). To be sure an isolated brain would be unable to learn anything; moreover it

would not even function normally, as shown by the experiments on sensory deprivation. However, looking, listening, sniffing and touching are not enough to take cognizance of anything beyond appearances. Recall that humans have always reproduced, yet did not discover the reproduction mechanism until a couple of centuries ago; they have presumably been speaking for many thousands of years, yet linguistics was not born until the 19th century; they have always lived in the midst of a gravitational field, but this fact was not discovered until recently; and they have been playing games of chance for millennia, yet the earliest ideas about chance and probability were not conceived until the 17th century. In short, sense experience and action are poor teachers: most people do not learn readily from either. (It is easier to learn to perceive and do.)

This difficulty to learn from sensing and doing has multiple sources: (*a*) appearance, unlike reality, is poorly structured; (*b*) all signals are embedded in noise; (*c*) perception and memory are highly selective: whereas at times they discard details, at other times they pass over essentials; (*d*) motivation is often lacking; (*e*) learning much and fast requires lots of knowing; (*f*) we learn most easily those items that happen to be consistent with our belief system, and tend to ignore those which are dissonant—as is the case with unfavorable evidence for our pet beliefs; (*g*) we are seldom equipped with comprehensive and coherent theories helping us "make sense" of experienced events, so we have nowhere to "file" the latter; (*h*) knowledge need not make us more receptive: learning develops a "set" (*Einstellung*) or habit that makes if difficult for the subject to face new problems or situations; (*i*) we may be insufficiently trained in logic, scientific method, statistics, and experimental design—all of which help learn from experience; (*j*) we may need to appear in the right for social reasons, and may thus be driven to simulate that we have got the message, or that there was no message to be received.

Some philosophers, particularly Popper and his followers, believe that we learn only from our mistakes, i.e. by punishment. Experimental psychology has refuted this view, showing that most people manage to learn a lot and yet they disregard negative evidence and seldom seek it (Wason, 1960; Smedslund, 1963; Jenkins and Ward, 1965; Einhorn and Hogarth, 1978). Consider the following situation. A subject is instructed to take action A if a certain variable X lies below cutoff point X_c, and action B otherwise. Further, his decisions have outcomes of two kinds: success in case a certain variable Y falls below a certain cutoff point Y_c, and failure otherwise. The state space of Figure 1.11 can then be built. Experimental

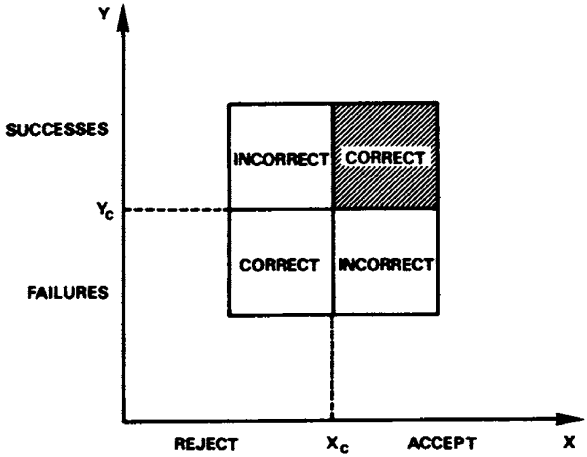


Fig. 1.11. A common fallacy. See text. Adapted from Einhorn and Hogarth (1978, p. 397).

studies (Einhorn and Hogarth, 1978) show that the only feedback most people care about is that coming from the pairs ⟨correct decision, success⟩, i.e. the points in the shaded cell. They generally ignore information in the three other cells, i.e. they tend to disregard the misses—as Bacon noted—and even the correct forecasts of failure (or non-occurrence). In other words, most people do *not* learn from their mistakes or even from their hits unless the latter bring rewards. Only a scientific training may instil the habit of seeking disconfirming information. In other words, ordinary experience is not self-correcting: it only gives us a chance to recognize error and learn to avoid it in future. Only scientific experience is consistently self-correcting.

If we knew more about learning we would not propose simple-minded epistemologies and would be able to learn in more efficient ways. As it is, we must make do with a handful of rules of thumb. One such rule is that *extremes of speed must be avoided in learning*: too slow a pace because it induces boredom, too swift a rhythm because it leaves no time for consolidation. We must regulate the inflow of information to keep it within moderate bounds. It must be quick enough to keep the neural systems alert to novelty: otherwise they become habituated or adapted, i.e. they do not respond. And the flow must not be so quick that psychons are dissolved no sooner they are formed, with the resulting confusion and poor recall. Low speed is best left to invertebrates, and high speed to computers: we had better choose medium speed and concentrate on selectivity, creativity, generality, and depth. Selectivity is becoming more and more important because we are subjected to an incessant bombardment with irrelevant

information: we must learn to evaluate and discard or skim most of the information that comes our way if we wish ever to make an original contribution to human knowledge. We must learn, in short, to choose wisely what to learn: "what we all need at this point in human evolution is to learn what it takes to learn what we should learn—and learn it" (Aurelio Peccei).

A second rule for efficient learning is: *Avoid cramming huge piles of unrelated data* ("facts"), for they are difficult to comprehend and recall. We tend to organize the stream of incoming information into "chunks" that can be understood and recalled as units or systems (George A. Miller). Thus coherent stories can be understood and recalled more readily than isolated sentences, and the latter more easily than nonsense syllables. The keys to this are of course connectedness (systemicity) and significance (what signs "say" about what). If this finding of experimental psychology were better known, chemists, biologists, medical researchers, social scientists and others might lower their resistance against theorizing, for theories are systems of knowledge. But of course theories are not data packages and therefore they will never replace data, such as addresses, formulas of chemical compounds, or historical dates. Hence we must learn such items alongside theories—but learning the latter will ease the learning of the former, for theories can be used as filing systems where data can fit.

A third rule to be followed is this: *Don't miss the apprenticeship period and get hold of the best teachers you can.* All animals capable of learning pass through a period of apprenticeship in some skill or other—e.g. flying, singing, fighting, exploring the environment, earning a living and, in the case of primates, even copulating. One does not become a farmer or a master builder, a manager or a research chemist, just by reading books, auditing courses, or watching television programs: every craft is learned at the knee of some master. One reason is that know-how cannot be wholly transmitted except by example. Another is that the text-book and the research paper give condensed and tidy reports on a process that has been anything but brief and tidy. A third reason is the need for incessant feedback (correction and encouragement). Those who object to apprenticeship calling it paternalistic, coercive, or repressive, will never train any disciples except in the arts of barren revolt and misleading rhetoric.

However, stressing the need for apprenticeship should not be mistaken for praising brainwashing, or even for favoring protracted apprenticeships. It is true that humans are remarkably immature at birth, and that they have to learn to cope with a habitat made extraordinarily complex by their forebears. Still, it is questionable whether the current apprenticeship period

is wise, the more so since it tends to increase. Doctorates in science are obtained in the West on the average at 29 years of age, i.e. they take 23 years of schooling not counting preschool. The reasons for this long and costly apprenticeship period seem to be to make sure that the candidate (*a*) has mastered the skills of the previous generation—which is unnecessary and impossible anyway; (*b*) will stick to the traditional ways (rather than try radical novelties), and (*c*) will stay out of the market place, thus protecting his teachers and himself. The apprenticeship should not be so short as to leave the student ill prepared to face professional problems by himself; but it should not be so long as to dry up every innovative and rebellious impulse. We should start earlier and be more demanding; but we should demand less curricular uniformity, cram less stray data, and allow more freedom of search and criticism.

Finally, are there limits to what an individual can learn? Is every one of us, as information processing psychology claims, a limited capacity information processor subject to Shannon's law concerning the limited capacity of every channel (Miller, 1967)? The limited capacity thesis applies to telephone lines, computers, and other artificial information systems, but it has not been proved for humans. There are indications that it does not apply to our brains: (*a*) unlike artificial information processors, our brains are plastic and, in particular, they have the property of self-organizability; (*b*) every time we learn something we become better prepared to learn further items: learning is an autocatalytic process, not the filling of prefabricated book shelves; (*c*) with practice we can learn to do two or more tasks at once—e.g. driving and chatting, typing and listening to music (cf. Neisser, 1980). Of course there are limits, some biological (e.g. poor genetic endowment, loss of neural plasticity, short life span) and others social (e.g. poor cultural environment, few opportunities). But these limits are not fixed once and for all: our brains and societies are to a large extent self-made. We can push the limits to what we can learn by building a society where everybody can freely exert the right to learn. We shall return to this theme in Ch. 13, Section 4.

3. Development and Evolution

3.1. *Cognitive development*

Organisms develop from birth and biopopulations evolve over time. Hence there are three different though mutually complementary ways of studying

biofunctions, in particular cognitive abilities: synchronically, developmentally, and historically. The synchronic approach attempts to learn what and how an animal learns; the developmental studies the growth and decline of cognitive abilities: how they are formed at an early age, what sequences they constitute, etc.; and the historical or evolutionary approach studies the origin and subsequent evolution of cognitive functions. Clearly, the synchronic study must initially precede all others, for it is impossible to discover the development or the evolution of X unless one has some idea of what X is. But once one has learned something about the development or evolution of X, one can form a better idea of what X is—if only because X has developed and evolved from some predecessor that shares some properties.

All the components of our body are genetically programmed, but they are fashioned, within the genic constraints, by the way we live and in particular the things we do. This holds for the muscles: compare those of an athlete and a sedentary professor, or a beef eater and a tortilla eater. And it holds for the brain, as the neurons of the plastic components of the brain can sprout dendrites and make synaptic connections all the time—or lose them with disuse or age. The brain, like the muscles, develops the way it is nourished and used. Thus our entire body is to a large extent self-constructed—particularly so the brain, which ultimately selects what food and what information it is to feed on. We are born with the ability to form new neural assemblies or psychons, but which ones will be formed depends on both internal and environmental circumstances. A severe protein deficiency in infancy may result in low intelligence, and repeated exposures to antisocial behavior may ensue in the same. The genome determines only the potential: which potentialities are actualized depends on the way the psychons develop, and this development is molded by both the natural and the social environments.

The brain of the neonate is immature: it weighs only one-fourth of the adult brain, and its cortex performs few specific functions. In particular its cortical neurons are small, have few and short dendrites, and are hardly interconnected. No wonder the newborn human is slow, clumsy, mindless, and therefore helpless. As the child grows up its neurons grow, sprout dendrites, and make synaptic contacts with neighbors. What began looking as a sparsely populated forest ends up by looking like a tangled jungle: see Figure 1.12. The maturation process continues throughout life unless senility sets in.

The maturation of the brain is a process composed of five main threads:

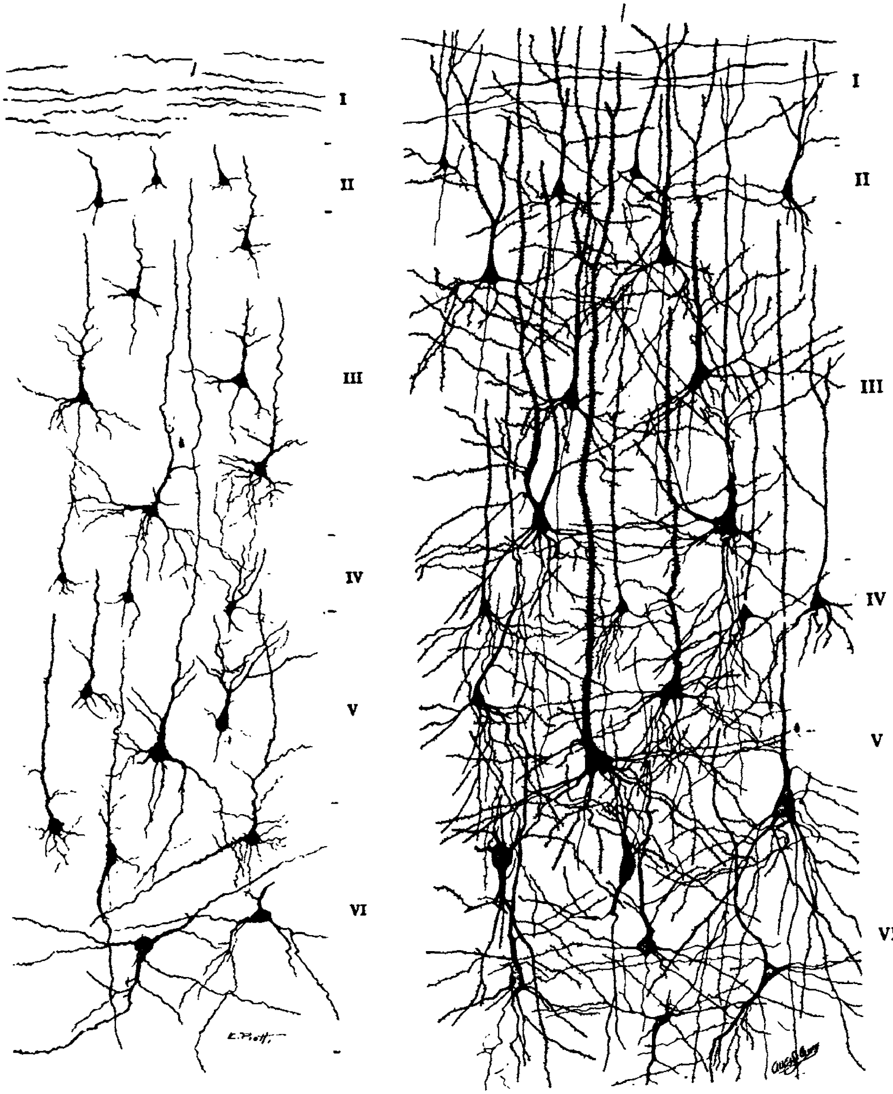


Fig. 1.12 Sections of the same region of the cerebral cortex of a newly born human (left) and of a 7 year old. From Conel (1939/67).

(*a*) growth of neurons; (*b*) growth of the myelin sheaths that wrap up and insulate the neuron axons; (*c*) change in the chemical composition of the neurons and the synaptic clefts; (*d*) proliferation of dendrites, and (*e*) multiplication of possible contacts with neighboring neurons, and thus formation of new psychons. Whereas processes (*a*) and (*b*) are completed in a few years, the others seem to go on for life. The maturation process is not homogeneous; in particular, the various cortical "areas" mature at different rates. Thus the motor "areas", which are partly prewired at birth, mature earliest; then come the sensory "areas", and in the first place the visual one. It is estimated that the cortex does not become fully functional until around 20 years of age. And it is known that the connectivity or interneuronal linkage remains changeable throughout life—barring senility. As the maturation process proceeds, new abilities emerge whereas others are lost. Thus it is likely that every one of Piaget's stages is the manifestation of the attainment of maturity (full functionality) of a large neuronal system.

The development of the nervous system can be somewhat accelerated by favorable environmental circumstances, and it can be distorted or even arrested altogether by unfavorable ones. For example, a kitten allowed to use a single eye during its first year of life will fail to develop binocular vision even after having been allowed to use both eyes (Wiesel and Hubel, 1965). A monkey shown only vertical lines in infancy will be unable to perceive horizontal bars later in life. A child who is not spoken to until age three may never learn to speak properly. In short, if the proper interneuron connections are not made in time, they are never made. In particular, certain cognitive abilities are not acquired or developed if the sensitive or critical period is missed. Practical consequence: Expose children and youngsters to rich and variable environments, and give them a chance to learn skills as early as possible. *Early training* is the ticket.

Developmental neuroscience and psychology, though still in their infancy, are in a position to solve the old nature/nurture (or nativism/empiricism) controversy, i.e. the question whether our cognitive abilities are inherited or learned. (The current dispute over the inheritability of intelligence is of course part of that controversy.) However, before rushing to take sides, or to propose a third alternative, we should clarify the rather vague notion of innateness.

According to contemporary biology "innate" subsumes two different concepts: "inborn" and "inherited though not present at birth". An inborn trait is one possessed at birth, such as the capacity to breath, move, cry,

look, or listen to spoken words in the case of human infants. Most inborn traits are inherited but some are outcomes of genic changes (e.g. mutations) or of original embryogenic processes. An inherited trait is one handed down by the animal's ancestors through the DNA molecules contained in the germ cells. Not all inherited traits are inborn; many emerge (or are "expressed") during development until the adult stage is reached. And, whether present at birth (inborn) or expressed later on (inherited), every such trait is ultimately controlled by the hereditary equipment (genome). In particular, an innate (inborn or inherited) cognitive function is one performed by a genetically programmed system. Thus "Animal *A* possesses innate capacity *B*" is spelled out "Animal *A* possesses neural system *C* which, when activated, discharges function *B*".

Now, like any other material system, from atom to society, a neural system has properties of two kinds: actual and potential. I.e. there are some "things" it does and others that it may do. Thus an atom is always moving relative to some thing or other (actual property), and it has the possibility of being excited by incoming radiation of certain wavelengths or, if already in an excited state, of decaying to some lower energy level. Likewise a neural system has, in addition to some permanent properties, others it can acquire, such as learning to perform certain functions. Indeed, whereas some neural systems are prewired (or develop into rigidly wired systems), others are plastic, i.e. they are formed, strengthened, or weakened in the course of life. (Recall Section 1.1.) Thus an infant starts learning the moment it is born: this shows that the ability to learn certain skills is inborn. On the other hand, in order to learn other skills the animal must wait for its nervous system to mature, and for the environment to offer the opportunity. However, this opportunity may never arrive, particularly in conditions of extreme biological and cultural deprivation. Suffice it to recall that children raised with insufficient food, parental care, perceptual and mental stimulation, and interaction with their peers, do not develop a normal brain: they can seldom if ever become intelligent.

In the light of the preceding we realize that nativism and empiricism, the psychological trends dealing with the nature/nurture controversy, are mutually complementary rather than exclusive (Dodwell, 1975)—and yet they are not exhaustive because neither acknowledges creativity. The reason both are needed is that, in the light of psychobiology, each is concerned with a certain kind of neural system. Nativism is (tacitly or explicitly) concerned with wired-in or rigid connections, which are typically innate (either inborn or formed later under the control of the genome). On

the other hand empiricism is (tacitly or explicitly) concerned with plastic or variable connections, which are typically acquired (learned).

Since the brain of a higher vertebrate (mammal or bird) contains systems of both kinds, we need empiricism as well as nativism—or rather a merger of the two enriched with the hypothesis of creativity, or the capacity of learning beyond experience. Only such a doctrine could account for the fact that we are born with a central nervous system that is partly rigid (prewired or programmed) and partly plastic (acquired or learned). But this is not all: although our plastic neural systems are partly self-constructed (by learning) they also have inherited capacities. Just as no two animals have exactly the same set of proteins—hence exactly the same chemical reactions—so no two animals are likely to have exactly the same neural systems and therefore exactly the same cognitive abilities. Such inherited neural differences, together with differences in environment and formal education, explain why some children are better than others at games or at languages, at mathematics or at drawing.

(A recent review of 111 studies published in the world literature, and covering 113,942 pairings, confirms the so-called polygenic inheritance of intelligence hypothesis, according to which the higher the fraction of genes two family members have in common, the higher the average correlation between their IQ's (Bouchard and McGue, 1981). However, the data exhibit also strong environmental influences. Thus whereas the average correlation between monozygotic ("identical") twins reared together is 0.86, that of monozygotic twins reared apart is 0.72; the corresponding figures for siblings reared together and apart are 0.47 and 0.24 respectively. Surprisingly, the figures for nonbiological sibling pairs are: 0.29 for adopted–natural pairs, and 0.34 for adopted–adopted pairs. Also, there is no evidence for the role of sex on general cognitive development. Finally, according to another study (Loehlin and Nichols, 1976), "identical" twins correlate only about 0.20 higher than fraternal twins in personality traits.)

In sum, the nature/nurture disjunction is not exclusive: genes determine only potentialities that are actualized or thwarted by the environment. What about innate ideas: are there any? Was Socrates right in holding that there is only remembrance, not acquisition of new knowledge? Before discussing this matter we should clarify the meaning of "idea", the more so since in the philosophical literature instincts and perceptions have sometimes been lumped together with images and abstract ideas. For the sake of precision we shall restrict ideas to concepts (e.g. "wet") and propositions (e.g. "The ground is wet."). The problem can now be restated as follows:

Are any concepts or propositions inborn or inherited? Innatism replies in the affirmative, empiricism in the negative. This 17th century debate has been reactivated in recent times in connection with the psycholinguistic problem of language acquisition—of which more in Ch. 3, Section 3.2.

The controversy about innate ideas is bound to remain unresolved as long as it is conducted on the superficial level of traditional mentalistic psychology. Neuropsychology (or psychobiology) is in a position to make a decisive contribution by suggesting that we study the language cortex and its function from birth to school age. We still know very little about this subject: psycholinguists seem to have been busier speculating and polemicizing than studying the child's brain and social environment. However, we do know this much: (*a*) ideation is an activity of fairly large neuron assemblies (psychons), not of germ cells, let alone of their DNA content, which is the only thing that gets transmitted from one generation to the next; (*b*) the brain of the newborn is so poorly organized (recall Figure 1.12) that it cannot even coordinate its movements, so that it is unlikely that it can form any concepts or propositions; (*c*) development is not merely the unpacking or "expression" of genic destiny: it is a creative process punctuated by the emergence of new functions and the exercise of some such functions in the acquisition of new knowledge; (*d*) children who grow up in culturally deprived environments have poor sensory-motor skills and a very poor inner life: although ideation is a brain function, it is not discharged in a social vacuum. In sum, we do not inherit knowledge but only the ability to acquire it in appropriate biological and social conditions. Which is just as well, for otherwise we would be lacking in curiosity, we would get stuck with ineradicable errors along with genetic defects, and the advancement of knowledge would have ceased long ago.

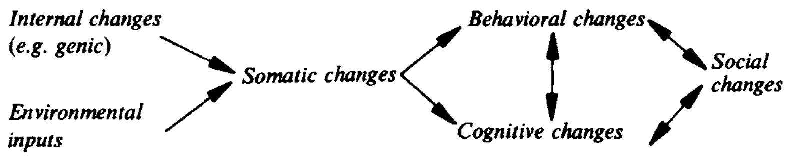
None of this entails that the newborn's brain is a clean slate ready to accept any inscriptions. On the contrary, it is partly a plastic brain with definite potentialities: it inherits not only certain rigid circuits (in charge of the basic household functions), but also the capacity to form new ones, as well as of dismantling some of them. Moreover we can form some such neural systems at will, as when we decide to learn a new motor skill, a new language, or a new branch of knowledge. In sum, we are unlikely to be born with any ideas, but we are born with the possibility of learning and even creating new ideas. (Maudsley (1876, p. 278) wrote that the notion of an innate idea is as absurd as that of innate pregnancy.) Shorter: although we are born ignorant we possess the neural apparatus to learn and, given the chance, we do learn quickly almost anything. Of course there are individual

differences traceable to genic differences: some people learn music, philosophy, or engineering, more easily than others. But this is beside the point: the point is that all such skills must be learned.

We share the instinct to learn with numerous animal species. However, we have a definite advantage over them, namely that of being born into a culture built by our ancestors. This inheritance spares us the need to make a fresh start in every single respect no sooner we learn to walk. (One's school record does not recapitulate the history of mankind.) But every cultural inheritance is ambivalent: it assuages one man's thirst for knowledge and incites another's exploration of new territory. Besides, whereas some societies encourage learning, others inhibit it, or at least they discourage inquiring into certain subjects. This being so, we will not learn much about learning if we restrict our inquiry to individual brains: we must study also the social matrix of the learner. More on this in Ch. 3, Section 2.

3.2 *Cognitive Evolution*

The evolution of cognitive abilities is an aspect of biological evolution and social history. All evolution proceeds by internal (in particular genic) variation and environmental (natural or social) pressure. Both factors may be pictured as inputs on the organism. In the case of animals the outputs of such externally induced or modulated changes are modifications of various kinds. Two such kinds of change are of paramount importance for the student of cognition: behavior (in the narrow sense of outcome of muscular activity) and cognition. These outputs are interdependent: new motor capabilities open up new possibilities of perception and ideation, and conversely. And both outcomes, the behavioral and the cognitive, are molded by society and in turn contribute to molding the latter.



The study of the evolution of cognitive abilities is of course one of the tasks of evolutionary psychology. Unfortunately the latter is still underdeveloped to the point of being almost exclusively speculative. There are two main factors of underdevelopment. One is the scarcity of data, almost all of which consist in circumstantial evidence. The other obstacle to the progress of evolutionary psychology is psychoneural dualism: by detaching mind

(or "software") from brain (or "hardware"), one forgets that what behaves or ideates is always some animal or other, and that animals happen to be subject to evolution. Nevertheless palaeoanthropologists and primatologists are beginning to learn a good deal about the evolution of the cognitive abilities of our ancestors over a time period of nearly three million years. (See Jerison, 1973; Masterton *et al.*, 1976a, b.) Such problems and findings should be of interest to all biologically oriented psychologists. By the same token they are a matter of indifference to mentalist psychologists, in particular those of the information processing variety, who claim to study either immaterial (disembodied) minds or computer programs, neither of which are subject to organic evolution.

Evolutionary psychology, then, is an underdeveloped branch of evolutionary biology. However, we cannot treat the evolution of the higher vertebrates in the same manner as that of the lower vertebrates, let alone that of invertebrates. Indeed the search for knowledge and possession of it modifies behavior, and behavior is not only an outcome of evolution but also a motor of it (Piaget, 1976b). In other words, by generating and using new knowledge the higher vertebrates can improve their survival strategies and thus modify the direction and pace of their own evolution. This being so, it is not true that for them, as for the lower animals, all variation is blind, and selection the only source of fitness—as assumed, e.g. by Campbell (1977). To some extent mammals and birds can control variations in themselves as well as in their environment, thereby improving their own fitness. Still, this does not entail that higher vertebrate (and particularly human) evolution is Lamarckian rather than Darwinian (as claimed by Piaget (1976b)). DNA, which is what we inherit, does not learn from experience. (For further criticisms see Bunge (1981a).)

A number of thinkers, notably Spencer, Helmholtz, Peirce, Mach, and Popper, have drawn a parallel between the history of ideas and biological evolution: in both cases there would be problems, trials and errors, and only the useful or adaptive novelties would "survive". This analogy passes sometimes for evolutionary epistemology, but it is not. To begin with it is a metaphor, not a theory. Secondly, it is a superficial analogy for, unlike organisms, ideas are neither alive nor self-existent, hence they do not evolve by themselves. New knowledge is not gained by random mutation, and scientific and technological knowledge is seldom gained by blind trial and error. Most new knowledge is obtained by inquiry, and there is no such search unless the animal is motivated (by need or by curiosity) and unless it knows some search strategies, however primitive.

A third objection to what passes for evolutionary epistemology among philosophers is that it is false that the evolution of knowledge consists simply in the elimination of "unfit" hypotheses, i.e. conjectures that fail to pass empirical tests. There is more to knowledge than hypotheses: there are also data and problems, viewpoints and methods, rules and instruments, designs and plans, etc. Fourthly, some patently false superstitions minted thousands of years ago—e.g. the doctrine that ideas dwell in a world of their own—are still going strong, whereas a number of scientific truths and useful technological innovations lie buried in the specialized literature, where they are likely to remain unless independently rediscovered or reinvented. Fifthly, truth does not always prevail once discovered or invented. It is sometimes suppressed for being incompatible with received dogma or vested interest, and at other times it is forgotten for lack of interest.

In short, most of what passes for evolutionary epistemology among philosophers is as removed from biology as it is from the history of knowledge. (For further criticisms see Currie (1978).) Genuine evolutionary epistemology takes organic evolution seriously, deals with the evolution of cognitive abilities as an aspect of brain evolution, and takes the social matrix into account. (For a sketch see Vollmer (1975).) True, evolutionary epistemology is so far little more than a research plan, so the philosopher cannot learn much from it. However, he should adopt a genuine evolutionary viewpoint, which involves dealing with concrete organisms immersed in concrete environments, rather than with disembodied minds (or computer programs) hanging in a natural and social vacuum.

4. Concluding remarks

It is legitimate to *distinguish* knowledge from the cognitive process. But it is impossible to *separate* them or to try to understand the one without the other—as impossible as to understand ashes without fire. Philosophy may well leave the detailed study of cognition to the cognitive sciences, but it cannot ignore them, and it must make such studies the firm (though changing) root of epistemology. It can disregard them only at the risk of producing obsolete and barren armchair psychology.

There are several approaches to cognition. One is to treat it as a particular kind of behavior, and to describe the latter in terms of stimuli and responses. We reject this, the behaviorist approach, because it ignores

that which does the cognition, namely the nervous system, and it does not even pose the problem of accounting for intelligent behavior. A second approach consists in begging the question: in taking ideas for granted and trying to account everything else in terms of them. We reject this, the mentalist approach, for being superficial and nonscientific: it fails to explain the nature of ideas, it assumes that they are self-existent, and it ignores the brain, thus isolating psychology and epistemology from all the other sciences. A third approach to cognition is the information processing one, i.e. cognitivism, which does not distinguish between brains and computers, is interested in the processing of information but not in its creation, and identifies body with hardware and mind with software. We reject this approach as well for failing to explain how information is generated and how knowledge guides (or misguides) behavior; it ignores the peculiarities of the nervous system (which it views as one more information processing system), as well as development and evolution, and it presupposes and therefore encourages psychophysical dualism.

The fourth approach to cognition is the *biosociological* one, the basic thesis of which is that every cognitive activity is a neural process, and one that interacts vigorously with other processes in the animal as well as with its natural and social environment. We adopt this approach for the following reasons (Bunge, 1980a, 1981a). First, the biosociological approach to cognition is in harmony with the ontology (or world view) of science, which is one of lawfully changing things rather than one of immaterial entities or disembodied processes. Second, it jibes with the life sciences, in particular (*a*) developmental and evolutionary psychology, which teach us that organisms change considerably in the course of their lives, and biopopulations may change to the point of speciation or extinction, and (*b*) neuroscience, which among other things can explain how the nervous system perceives, imagines, feels, thinks, etc.—or fails to discharge satisfactorily any such functions. Third, the biosociological approach to cognition harmonizes with the social sciences, in particular the sociology of knowledge, which shows that society molds the way we see the world and ourselves, and studies the ways in which society encourages or discourages inquiry, and benefits or suffers from its results.

CHAPTER 2

KNOWLEDGE

We may distinguish knowledge from the process of its acquisition: we may think of knowledge as a product or outcome of cognitive operations such as perception and inference. But of course there is no outcome detachable from process: the former is just the final phase of a sequence of events. Nor is knowledge the invariable outcome of a cognitive process: sometimes inquiry ends up in ignorance, at best in the resolve to engage in further inquiry. In any event knowledge, when acquired, is a collection of learned items: it is a collection of brain processes or a disposition to replay them. Yet although we cannot detach the outcome (knowledge) from the corresponding process (cognition), we may distinguish them.

Moreover for some purposes we may *feign* that cognitive processes have a "content" that can be communicated to other brains or externalized as artifacts such as inscriptions and tapes. To be sure actually there is no such content and, a fortiori, no such transfer. Acquiring knowledge is learning something, i.e. going through a certain brain process, hence not the same as acquiring a book or some other commodity. Likewise exchanging information is not like trading things but is interacting with another animal in such a way that each party elicits certain learning processes in the other's brain.

Yet from a methodological point of view it is convenient to feign that cognitive processes do have a transferable content, so that we may think of the latter separately from the former. This convenient fiction amounts to supposing that, for certain purposes—such as checking validity or usefulness—it does not matter what or who went through the cognitive process in question. It could have been Archimedes in Syracuse two millennia ago, or Einstein in Bern at the beginning of our century. This distinction between cognition and its content is useful in epistemology and there is no harm in it as long as it be understood as a fiction, not as an ontological thesis on the autonomy of ideas (Bunge, 1981a). After all, mathematics too deals with fictions—though not wild or idle ones—and this does not detract from its usefulness. On the contrary, the distinction guarantees impersonality and facilitates universality.

We shall also distinguish knowledge from belief, but this distinction is

not a mere methodological gimmick. That knowledge is not belief should be clear from reflection on cases such as "A knows B but does not believe B", and "A believes B although he does not really know B". Yet there is a large body of philosophical literature that defines knowledge as a kind of belief; we shall pay hardly any attention to it.

Finally we shall sketch the features of the deliberate and systematic acquisition of knowledge through inquiry, leaving its detailed study to later chapters. And we shall characterize a field of inquiry in terms of a material framework and a conceptual one.

1. From Cognition to Knowledge

1.1. *Knowledge Acquisition and Use*

We have assumed that every cognitive activity is the activity of some plastic neural system or psychon (Ch. 1, Section 2). We must now distinguish between acquiring (learning) a piece of knowledge and using it (e.g. recalling it or putting it into practice). From our neuropsychological point of view this distinction amounts to distinguishing the *original formation* of new psychons (or psychons of new types) from the *activation* of existing psychons (or the formation of psychons of a type that already had representatives in the individual's brain). In other words, we distinguish between the assembly of new neural systems (or circuit rewiring, in the old metaphor) and the use of existing ones. Presumably the lower animals can do only the latter: they do not self-assemble or self-organize new neural systems, except perhaps through malfunction. But we do it as long as we learn new items.

Example 1: An animal is in the process of learning to pair stimuli of kind E to responses of kind R, each accompanied by internal states belonging to the brain state space S. (For the notion of a state space see Ch. 1, Section 1.2.) That is, the brain of the animal is establishing a new psychon "embodying" (represented by) a state function of type $F{:}E \times S \rightarrow R \times S$. Note that this is a very strong assumption, for it implies that a finite number of stimuli can elicit the emergence of a property represented by a function F from $E \times S$, which may contain infinitely many elements, to $R \times S$, which may also be composed of infinitely many members. Clearly, this conjecture goes against the grain of inductivism, which demands that we learn item by item and synthesize only at the end. (More in Ch. 5, Section 2.1.) And the

hypothesis is equally alien to innatism: in fact it is an alternative to both rationalism and empiricism.

Example 2: The animal has already acquired the function F in the former example, and so it has acquired the *disposition* to make responses of type R every time it receives stimuli of type E. It is now subjected to a stimulus of this kind (though possibly different from any of the individual stimuli that induced its learning F), and it generates both the suitable response and the accompanying inner state. It is making use of a bit of knowledge acquired earlier, namely the one summarized by the function F.

To give some precision to the preceding ideas we introduce a certain notion of functional equivalence. We say that the linear functions $y = x$ and $y = ax + b$ are equivalent, or belong to the equivalence class of linear functions. Likewise the trigonometric functions $y = \sin x$ and $y = \sin (ax + b)$ will be said to belong to the equivalence class of sine functions. In general, we say that two functions differing only by a linear transformation of their independent variable(s) are *equivalent*. And we write $[F]$ for the *equivalence class of* F, i.e. the set of all functions equivalent to F. We also say that, if the state functions of two concrete (material) systems are equivalent, then the systems themselves and their activities (i.e. the processes occurring in them) are *functionally equivalent* no matter how much they may differ in details.

We are now ready for our definitions of knowledge acquisition and use. Let b be an animal endowed with a plastic neural supersystem P, and consider a time interval τ. We stipulate that

(i) *b acquires knowledge of kind* $[F]$ over period τ if and only if a new psychon, characterized by a state function F (a member of $[F]$), is formed in P in the course of τ;

(ii) *b makes use of knowledge of kind* $[F]$ over period τ if and only if the P of b contains at least one psychon, active during τ, characterized by a state function of kind $[F]$ at the beginning of τ;

(iii) the *total knowledge* of b at time t is the union of all the bits of knowledge, of all types, that b has acquired up until t. We call this set $K(b, t)$.

Note that, contrary to the behaviorist tradition, but in consonance with cognitivism, we have equated learning with the acquisition of knowledge (see, e.g. Dickinson, 1980). However, in contrast to mentalistic cognitivism, we equate in turn the acquisition of knowledge with a partial restructuring of the nervous system rather than with the formation of a new immaterial mental configuration. Our hypothesis is more precise and

therefore better testable than the mentalistic one, which contains the essentially vague notion of a mental configuration. Moreover our hypothesis has recently received a sensational experimental confirmation from the study of bird song. In fact, it has been found that in the spring, when male canaries learn new song repertoires, their song control neural nuclei nearly double in volume relative to the fall, when they stop singing (Nottebohm, 1981). This seems to indicate that learning is (identical with) the growth of new dendritic segments and the consequent formation of new synapses, whereas forgetting is the shedding of synapses.

Note also that our definition allows for considerable variability among individuals, and even in the course of the lifetime of any given individual. Thus even if two individuals learn "the same thing" (e.g. to shake hands, or to read graphs), there may be considerable differences in the details of their learning processes. Consequently there may be differences in the outcomes or performances, e.g. in the way they shake hands or read graphs. Second and more important, learning is not done by the mythical hollow organism of behaviorism or by the mythical disembodied mind of the mentalist: learning consists in some nervous system acquiring new properties. (Other organs, such as the heart and the kidney, do not have this capacity: they have fixed state functions.) Third, there is no instantaneous learning, hence no instant knowledge: getting to know anything takes time, however short.

Our definitions have a number of consequences, some of which are irrelevant to traditional epistemology whereas others are incompatible with it. Here are some of them:

(i) An animal lacking any plastic neural systems does not know anything. (Thus a worm has no more knowledge than a stone. When it learns to withdraw from a noxious stimulus, all it does is to inhibit or block certain neural pathways: it does not create any new ones.)

(ii) An animal whose plastic neural system fails to contain a subsystem characterizable by a state function of kind $[F]$, and active during τ, has no knowledge of kind $[F]$ during τ. (Inactivity may be due to sleep or coma, or it may be caused by drugs.)

(iii) All cognitive deficits and abnormalities are dysfunctions, temporary or permanent, of the corresponding neural systems. (Such dysfunctions may occur at the molecular level, as is the case with excesses or deficiencies of neurotransmitters, or at the cellular level, as is the case with "wrong" synaptic connections.)

Acquiring knowledge and using it are cognitive activities, but not every cognitive activity is successful: we may conduct an investigation without

learning anything. In other words, the collection of learning activities is properly included in that of cognitive activities. Consider the collection of all the learning activities of an animal b up until time t, i.e. the total knowledge $K(b, t)$ of b until t. We obtain the knowledge of the species to which b belongs by forming the union of all the individuals' knowledge of the species. Thus what we call *human knowledge* at a given moment is the union of all the bits of knowledge—correct or incorrect—acquired by members of the human species up until that moment, i.e. $K(H,t) = \bigcup_{b \in H} K(b,t)$. This is a large collection, for it includes every truth, half truth, and error that every single human has ever learned. We risk nothing by postulating that nobody will ever learn all of human knowledge. Which is just as well, for much of it is not worth acquiring anyway.

For most of prehistory knowledge existed only in individual brains. With the invention of drawing, painting, sculpting, and particularly writing, knowledge could be encoded and externalized in *cultural artifacts* that could circulate throughout the community. This facilitated the storing, sharing, and enriching of knowledge. But it also fostered the myth of the independent "content" of knowledge—independent, that is, from the inquiring subject.

It is easy to see how this myth can be generated and maintained. When somebody finishes a paper or a drawing, these pieces of matter can be detached and seen by somebody else: even their creator can stand back and contemplate them as if they were self-existing, while in fact their "content" depends on their being perceived and understood by some brain. This creates the illusion that we are in the presence of three separate items: the neural (and motor) process resulting in the writing or drawing, the cultural artifact, and the knowledge or feeling encoded in the latter. The next step is to collect all such bits of knowledge detached from brains—i.e. all the problems and data, theories and plans "in themselves"—and endow such a collection with a life of its own. The final step is to give a name to such a collection of items allegedly hovering above brains and society—e.g. the "realm of ideas", the "objective spirit", or "world 3". (See, e.g. Plato, Hegel, Bolzano, or the later Popper.) In this way the illusion is created that such "worlds" of knowledge and feeling persist or subsist once they have been created by concrete individuals in specific social circumstances, and moreover that they can interact with living beings. Actually those collections do not constitute worlds (systems), for they are quite heterogeneous, particularly if cultural artifacts are counted among their inmates. Worse, there can be no empirical evidence for the hypothesis that such ideal

"worlds" lead an existence separate from living brains. (Cf. Bunge, 1981a.)

To return to learning. The statement that somebody has learned a bit of knowledge is ambiguous. It may signify that she has created some new knowledge, or that she has grasped a bit of "existing" knowledge—existing, that is, as a process in someone else's brain or encoded in some cultural artifact such as a book or a film. The difference is that between production and consumption, and it is both a psychological and a social difference. It is psychological because producing new knowledge is generating psychons of a type that had not existed before in the evolution of the species, whereas to learn a bit of existing knowledge is to generate a psychon similar to one that had already been activated in someone else's brain. The difference is also social, for 'original knowledge' means "knowledge not available to anyone else before".

A bit of knowledge, whether original or unoriginal, may be thorough or partial, and it may be of high or low grade. Thus we may have partial or full knowledge of a gossip or of a theorem. So learning, whether afresh or second hand, comes in degrees. (Moreover we shall postulate that only constructs may be fully known. But this must await Section 5.) Unfortunately we still do not have a measure of degrees of knowledge.

As for quality, recall the (unclear) difference between knowledge and wisdom. The wise person is eager to learn discriminately. Knowing (having learned) much is not the same as being wise. A person may be a mine of information, yet quite incapable of creating new knowledge or correctly evaluating knowledge produced by others. Unlike erudition, wisdom is quality knowledge: it is knowing important things and knowing how to find out more of the same. Beware of those who claim that wisdom is disjoint from, and superior to, knowledge: they are likely to be superstition mongers. When we hear T. S. Eliot complaining "Where is the Wisdom we have lost in knowledge?", we should reply: "And where is the knowledge you parade as wisdom?"

Our view of knowledge differs radically from that of the ordinary language philosophers who hold that to know X is to know how to talk about X. Such knowledge would be "embodied" in certain speech rules rather than in bodies of data or in theories. Thus we are told that we may use the word 'quark' just in case there is a rule on how to use it. The stipulation has the advantage that it enables anyone with a command of ordinary language to pass for an expert on anything that can be talked about in the language.

Our view differs also from the empiricist doctrine according to which

every learning process is gradual and proceeds by trial and error. We have made no such assumption, although we did assume that every learning process takes time. For all we know a number of higher vertebrates can learn certain "things" at the first go. Besides, all learning is driven by motivation and expectation. Also, not all trials are tests: errors are not easily eliminated even after ending up repeatedly in disaster. Just think of war and prejudice. The weight of prejudice is such, that learning new ideas often requires getting rid of false preconceptions (Bacon's *idola*). Thus one of the main obstacles to the learning of mechanics is the old superstition that a moving body eventually comes to rest for exhaustion of a driving force, fuel, or impetus (McCloskey *et al.*, 1980). Likewise the science student who wishes to learn some philosophy must begin by criticizing two views he learned in his science textbooks, namely that all concepts must be defined operationally, and that theories are data summaries.

Finally, our view is at variance with the conception of "knowledge as a matter of conversation and of social practice, rather than as an attempt to mirror nature" (Rorty, 1979, p. 171). To be sure, communication disseminates knowledge and social practice checks it: there is no cognition in a social vacuum. Still, cognition is a brain process not a social one, and we must have some information before we can trade it in conversation. Likewise there is no respiration in a physical vacuum, but this does not prove that respiration is an atmospheric or an ecological process. Besides, what would be the point of inquiry if it did not result in a representation of nature (and society)? The view of knowledge as a matter of conversation and social practice may fit the sophist but is a caricature of the investigator, be he a scientist, a technologist, or a humanist.

1.2. *Knowledge Evolution*

Our qualitative representation of knowledge in the preceding section allows us to speak of the *state of knowledge* attained by an animal at a given time, namely by identifying such state with the total knowledge acquired and retained by the animal up until that time. We denoted such knowledge by $K(b, t)$ (Section 1.1).

This representation allows one to compare any two cognitive states, be it of the same animal or of different animals at different times. Such comparison may yield, in principle, either of the following results. One of the animals knows *more* than the other (or the same animal at a different time), the two *share* some knowledge though not all, and the two do not

share any knowledge. The first case might be that of the lower animals capable of learning something; the second, that of all primates; and the third pigeonhole seems to be empty. We assume then that *any two humans share some knowledge* and that *every one of them knows something that the other does not*. Without some common knowledge we could not communicate with one another; without some idiosyncratic knowledge there would be no point in communicating.

(The comparison of cognitive states can be synchronic or diachronic. In particular it allows us to estimate the gain or loss of knowledge in the course of time. Calling $K(b, 1)$ the cognitive state of animal b at time 1, and $K(b, 2)$ that at time 2, we may say that the *knowledge gain* between times 1 and 2 is $\gamma_b(1,2) = K(b, 1) - K(b, 2) = K(b, 1) \cap \overline{K(b, 2)}$. The gain is null, i.e. $\gamma_b(1,2) = \varnothing$, if the animal has learned nothing during that period. What holds for individuals holds, *mutatis mutandis*, for groups, upon introducing the definition of group knowledge as the sum total (union) of the knowledge of its members. In obvious symbols, $\gamma_G(1, 2) = K(G, 1) - K(G, 2)$, where $K(G, t) = \bigcup_{b \in G} K(b, t)$.)

The same formal tools allow us to compare (in principle) the knowledge states of different communities, or even of mankind at different times. If we do so we may find that there are always gains and losses. Thus we know many "things" that our ancestors ignored, but we also ignore much that made the existence of our forebears possible (and of others that made it unnecessarily harsh). In any case individual knowledge develops and human knowledge evolves: the former normally grows, the latter grows over certain periods but not over others. (Caution: knowledge can neither develop nor evolve, for it is not a concrete thing. Only individuals develop and communities evolve. However, the expressions 'development of knowledge' and 'evolution of knowledge' are permissible as long as they are understood to be elliptical.)

There are several views on the evolution of human knowledge. They can be grouped into two classes: those which hold, and those which deny, that the evolution of human knowledge follows definite patterns. Let us take a quick look at two well-known members of the former class. The best known of them is dialectics, according to which every construct carries within itself the seeds of its own negation, so eventually turns automatically into its opposite. This opposite would at the same time cancel and preserve the previous stage, and would eventually be subject to the same fate, i.e. negation. This second negation, far from being identical to the first construct (or thesis), would synthesize the first and the second: it would

contain whatever was valuable in them, so it would occupy a higher rung in the ladder of progress. (Do not ask for more precision or for the mechanism whereby one item gets transmogrified into its opposite: dialecticians are neither clear nor thorough. The doctrine was proposed several millennia ago and has never gone beyond the verbal and metaphorical stage.)

The dialectical schema of the evolution of ideas is attractive at first blush: it is simple, it promises continuous progress, and it can be exemplified provided one is not too exacting concerning the notion of dialectical negation. However, the doctrine cannot be taken literally, for the only clear case of negation of an idea, namely logical negation, is not dialectical. In fact a proposition and its negation do not join to form a higher synthesis but only either a worthless contradiction or a harmless tautology. Also, it is an idealistic thesis, for it concerns the development of ideas in themselves (i.e. nonembodied ideas) rather than that of brains or communities. Thirdly, dialectics is confirmed by some examples but refuted by others. Think of the many futile controversies—theological, political and philosophical—that have occupied so many good brains for so many millennia without ever yielding any new, let along superior, knowledge. Or think of the way the great scientific theories have been created. What was the dialectical movement (of ideas) that ensued in Archimedes' statics, or in Newton's dynamics, or in Darwin's theory of evolution, or in Maxwell's electrodynamics, or in quantum mechanics? No dialectician has taken the trouble of telling us. In conclusion, there may be examples of dialectical process in the evolution of knowledge but there are also counter-examples, so there can be no general patterns or laws of cognitive dialectics. And, since there are no such laws, the principles of dialectics fail to explain or predict the evolution of knowledge. (More on dialectics in Bunge (1981a).)

The same holds for Popper's generalization of dialectics, i.e. what he calls the "method" of trial and error, summarized in the sequence: problem-theory-criticism (conceptual or empirical)—new better theory (Popper, 1972). For one thing this conception too is idealistic, for it deals with ideas in themselves in isolation from brains and societies. For another, not every theory, even if accepted, need be better than the theory it replaces: there are examples of retrogression or involution in the history of ideas. Thirdly, Popper's schema, which is a sketch of the usual conception of the scientific method, is normative not descriptive: it prescribes the way to go about investigating anything, but does not describe the actual cognitive evolution of mankind, or even of modern science. Hence it cannot

be regarded as a generalization of the "laws" of dialectics. So much for two popular views on the evolution of knowledge.

It is doubtful that anybody has found *laws* of the evolution of knowledge. (See however Wilder (1981) for a set of alleged laws of the evolution of mathematics.) Consequently we do not even know whether there *are* any such laws. But we do know that, because ideas are not self-moving, we will not find any laws by examining ideas in themselves: we must investigate instead cognitive systems, i.e. learning individuals and communities, rather than the nonembodied "product" of their inquiry. This shift from ideas to their creators displaces the entire problem: it has now become a subproblem of the wider but more tractable problem of the laws of history. If there are such laws then the pattern of the evolution of knowledge is just an aspect of such laws. But, again, thus far we do not know whether there are laws of history. And in any case we must postpone this intriguing question to Vol. 6.

The result so far seems disappointing: we have succeeded only in proposing a qualitative representation of the state of knowledge and of cognitive change, and have expressed skepticism concerning the very existence of laws of such change. There are alternative approaches, two of which are quantitative. Let us see what they are and how they fare. The simplest of them is just to count knowledge items. This is not even a rough measure of knowledge, for the following reasons. One is that a mnemonist can memorize a huge amount of trivial information, whereas a foundational worker may be able to learn only a few but powerful items. Another, even worse difficulty, is the question of what is to count as a unit of knowledge: a proposition, a list of propositions (such as the one in a telephone directory), or a theory containing infinitely many propositions? In short, sheer numbers will not do.

A second and far more popular proposal is that of the subjectivist or Bayesian (or personalist) school of probability. It claims that (*a*) knowledge is a kind of belief, (*b*) belief comes in different strengths, (*c*) such strengths are measured by probabilities, and (*d*) probabilities are nothing else but degrees or strengths of beliefs, and so changes in probability values represent always changes in belief states. The centerpiece of the Bayesian school is the subjective interpretation of Bayes' theorem, which gives the probability of a hypothesis in the light of an item of empirical evidence, in terms of the prior (a priori) probability of the hypothesis together with that of the evidence and the likelihood of the latter given the hypothesis.

There are several objections to the theses that probabilities represent

(only) states of knowledge or of ignorance, and the Bayes' theorem (or its information-theoretic rewrite) represents the learning process and even the evolution of human knowledge. Firstly, it leaves out problems and plans, methods and evaluations—in sum, whatever is neither hypothesis nor datum. Secondly, probabilities can be assigned to propositions only by decree: there are no objective criteria for such assignments since, according to the Bayesians, all probabilities are subjective or personal. Thirdly, a condition for the applicability of probability is chance or randomness—a condition absent from every fairly well organized body of knowledge. Finally, it is doubtful whether many people do learn from experience in any rational way, let alone in accordance with Bayes' theorem—particularly when probabilities are involved. In short, subjective probability does not solve the problem of measuring the state of knowledge nor, a fortiori, its evolution.

(Compare the claims of subjectivists with the actual way probabilities are used in science and technology. Here there are two possibilities: either one knows the probabilities of interest or one does not. If the former, it is either because of measurement or of computation, and in either case the posited probabilities are *objective* in the senses that they quantitate objective possibilities and that they are subject-free, i.e. public. If on the other hand the scientist or technologist is faced with total ignorance concerning the probabilities of alternative possibilities, or the distribution of a continuous stochastic variable, he adopts either of the following strategies. Either he abstains from assigning any values to the probabilities or the densities in question, or he makes hypothetical assignments, watches for the logical consequences—and makes the necessary corrections. He realizes that any such hypothetical assignments—e.g. equiprobability—are *hypothetical* and part of the particular model not of the mathematical theory of probability. On the other hand the Bayesian statistician makes arbitrary assignments of prior probabilities without regard to any matters of fact: he defies the scientific method. To the scientist or technologist any probabilities are probabilities of states or events, and therefore they (*a*) depend on the nature of the thing concerned and (*b*) must be either measured or conjectured regardless of the verdict of the subjectivist probabilist. More in Bunge (1981c).)

Finally, science and technology watchers and administrators have developed objective though ambiguous indicators of the volume and rate of growth of public knowledge. In fact there is an entire industry devoted to such measures as well as specialized periodicals, such as *Citation Index*, and

annals such as *Science Indicators*. No doubt, such "scientometric" indicators do measure the *volume* of publications and their *diffusion*. Whether they also measure the volume and growth of *knowledge* is doubtful. Merely counting publications does not discriminate between original research and rehash, between deep insight and shallow information, or between revolutionary proposal and routine stuff. And counting the number of times a publication is cited indicates popularity rather than intrinsic value. The difference is important for, as a rule, the newer and deeper a contribution, the harder it is to understand and therefore the least popular it is bound to be at least for a while.

2. Modes of Knowledge

2.1. *Basic Categories*

Classical epistemology dealt with the lone knowing subject and his knowledge, and did so in a static manner: it minimized the cognitive process and ignored its interaction with social behavior. The epistemology we are sketching deals instead with animals engaged in learning processes in their environment, and regards knowledge as a phase in those processes. Thus it interprets the conventional expression 'X is known' as "There is at least one animal of species S who has learned X under circumstances Y". At the same time we recognize the need for studying the "product" of the cognitive process regardless of the idiosyncrasies of the learning subject and her environment—i.e. the need for the study of knowledge.

Now, 'knowledge' is an omnibus word: it covers all learned skills (not the inherited ones though) and all manner of ideas. Indeed knowledge can be:

(i) *sensory-motor*—e.g. knowing how to walk (or limp) or how to type (however clumsily);

(ii) *perceptual*—e.g. tasting a lemon as bitter (or otherwise), or seeing a knoll of grass as green (or purple);

(iii) *conceptual* or *propositional*—e.g. knowing that the Earth revolves around the Sun (or conversely), or that the heart pumps the blood (or is the seat of the soul).

To put it into classical terms: knowledge can come from any of three "sources"—action, perception, or ideation. Therefore radical pragmatism ("Action is the only source of knowledge") does contain a grain of truth. So do empiricism ("Perception is the only source of knowledge") and rationalism ("Cogitation is the only source of knowledge"). By the same

token each of these epistemologies is mistaken in excluding the other two "sources" of knowledge. (Caution: memory, traditionally regarded as a source of knowledge—about the past—can at most distort acquired knowledge, never generate it.)

The above distinction among the three knowledge categories should not blind us to the fact that they are interrelated. Thus conceptual knowledge can improve motor skills and perception: the trained hand (or eye or ear) performs better than the naive one. Besides, all three are simultaneously present in many cognitive activities, such as drawing and writing.

Note also that knowledge of any type may be correct or incorrect to some degree. Thus most of us would be unable to make anatomical drawings like those of Ramón y Cajal, or even Vesalius or Leonardo; most city dwellers fail to perceive (sense and identify) the sounds of nature as acutely and precisely as country folks; and all of us know as many falsities as truths. (And this trivially, in the sense that whoever knows the proposition p knows *eo ipso* not-p, as well as in the deeper sense that much of our conceptual knowledge is at best approximately true.) In short, knowledge is not necessarily correct. Nor does knowledge involve belief: I know many "things" (e.g. ideological doctrines) in which I do not believe, and I grant beforehand that some of my beliefs are false. Therefore the standard definition of "knowledge" as true belief is inadequate. More in Section 3.2.

A second useful distinction is that between *self-knowledge* (or knowledge of oneself) and *other-knowledge* (or knowledge of something other than oneself). Both can be obtained directly through perception, and self-knowledge also from feeling and emotion. (Feeling and emotion are not cognitive operations, but our awareness of them is a cognitive item.) They can also be known indirectly, namely by cogitation. And either mode supplements, checks, and corrects the other.

The ancient and wise injunction *Know thyself* can now be rendered more precise, to read *Know thy brain and thy social behavior.* To be sure one will never get to know much about oneself without the help of neuroscientists, psychologists, and social scientists. However, one can learn a lot about oneself by watching one's feelings and desires, modes of thought and behavior patterns—all of which *is* part of knowing one's brain and one's social behavior. A scientific study of all that will yield more detail and, in particular, a deeper understanding of why we feel, desire, think and behave the way we do.

Other-knowledge can be of things (natural, social, or artificial) or of

constructs (concepts, propositions, theories, or proposals). And the latter may, though it need not, involve the former. Thus knowing a reasonably true atomic theory implies knowing something about atoms. On the other hand knowing some differential equations does not imply knowing any real things except the inscriptions representing such equations. (Rather on the contrary, we need some knowledge of differential equations in order to get to know something about atoms in any depth.)

We emphasize that knowing facts is not the same as knowing propositions about facts—a distinction that epistemic logic fails to draw. When perceiving fact f, or performing action f, we may gain *direct* knowledge of f; this knowledge is not necessarily propositional. Hence we are justified in asserting "I know f", rather than "I know that the proposition p describing fact f is true". On the other hand when learning about facts in books, or when working out scientific theories, we have *indirect* or propositional knowledge of them. These two modes of knowing, κ_1 and κ_2, compose as shown in the following self-explanatory diagram:

$$\begin{tikzcd}
\textit{Facts} \arrow[r, "\kappa_1"] \arrow[dr, swap, "\kappa_1 \circ \kappa_2"] & \textit{Propositions} \arrow[d, "\kappa_2"] \\
 & \textit{Propositions}
\end{tikzcd}$$

Table 2.1 exhibits some types of knowledge distinguished by subject matter or referent.

A third useful distinction is that between first and second hand knowledge. *First hand* knowledge is acquired by personal experience, in particular research. *Second hand* knowledge is knowledge about first hand knowledge: it is communicated by wòrd of mouth, the media, textbooks, or review articles. This partition of knowledge items does not involve any

TABLE 2.1
Types of knowledge with regard to subject matter or referent

Type	*Referent(s)*	*Field(s)*
Formal	Constructs	Logic, mathematics, semantics
Factual	Things, facts	Ordinary knowledge, arts, crafts, factual science, technology, ontology
Empirical	Experience	Ordinary knowledge, arts, crafts
Moral	Right conduct	Ethics, sociology, art
Epistemological	Knowledge	Epistemology

valuation. First hand knowledge may be worthless—e.g. I know there is a tiny inkspot on this sheet of paper—and second hand knowledge may be high grade—e.g. an insightful critical review of recent work on the effectiveness of psychotherapy. Knowledge of both types is necessary, for nobody can learn everything first hand. But we should not allow second hand knowledge to get the upper hand—as it did during the Middle Ages.

A related distinction is that between private and public knowledge. We may say that animal b has *private* (or privileged) knowledge of X if and only if there is nobody except b who has knowledge of X. Otherwise, i.e. if an item of knowledge is shared among at least some members of a society, it is *public* in that society. These are just definitions. Finding out whether a given bit of knowledge is public in some society calls for empirical investigation by psychologists, anthropologists, social psychologists, sociologists of knowledge, or historians of culture. Mere linguistic analysis will not suffice, notwithstanding those philosophers who claim that 'p is known' is an obvious example of public (or impersonal) knowledge (e.g. Williams, 1972). The fact that someone holds something to be known, without caring to mention who knows it, does not entail that such knowing can be impersonal or suprapersonal: it only indicates that the individual who makes that claim does not care to tell us who knows that particular item. The ordinary language expression 'p is known' must be analyzed into 'There is some animal that knows p'. Likewise when we say that 'x attracts (or repels)' we presuppose that there is something, other than x, that is attracted (or repelled) by x. *Moral*: Language is not a tool of analysis but an object of it.

There are two kinds of private and two kinds of public knowledge. Private knowledge can be knowledge of one's own states, in particular brain states, or secret knowledge, i.e. knowledge kept "classified". The former is direct in the sense that the subject has access to it without the help of anyone else. On the other hand getting hold of a secret concerning somebody or something requires data supplied by others—deliberately or unwittingly. However, knowledge of the two kinds can, in principle, be made accessible to others, i.e. rendered public in part. Surely your headache will not be felt by anyone else, but it may be guessed from your grimaces or from objective indicators such as, perhaps, the rate of endorphin synthesis in your brain. And, if physiological psychology continues to advance at its present pace, conceptual privacy may soon vanish. So much for the claim that all mental events are private.

Public knowledge too comes in two varieties: *ordinary* and *specialized.*

The former is shared by (nearly) all (adults) of a given community (at a given time), whereas specialized knowledge circulates (at a given time) only within very special learning communities, such as those of masons or mathematicians. The philosophers who possess no specialized knowledge tend to revere ordinary knowledge, claiming that it is sufficient for all purposes, or at least that it should be regarded as the Supreme Court of Cognitive Appeals, for being solidly anchored in experience and untainted by prejudice and theory. Both claims are wrong. The mere existence of mathematics, science, technology, and philosophy shows that common sense was found insufficient, for they start off where the latter stops or founders. Nor is ordinary knowledge as solid as it looks: it is mostly superficial, largely inaccurate, and remarkably incomplete. And in any event ordinary knowledge is shaped and updated not only by personal experience but also by the entire culture, as it feeds off the stale crumbs of science, technology and philosophy that may come within its reach. (By the way, there is no such thing as "common sense philosophy". This is a contradiction in terms: whatever is commonsensical is uneducated and therefore likely to be uncritical, shallow, and obsolete—in short, unphilosophical.)

The second variety of public knowledge, namely specialized knowledge, is supposed to be a result of specialized practice or of inquiry by individuals or teams possessing a special background and special skills. It comes in a variety of grades, from expertise (e.g. in cooking or managing or computing) to scientific, technological or humanistic knowledge. Here again, truth and usefulness are irrelevant: whereas some artisanal know-how is true and effective, some scientific (or technological or humanistic) knowledge is false and useless.

Plato's dichotomy between *doxa* (groundless opinion) and *episteme* (demonstrable science), though widely accepted, is inadequate. An electronic technician, though no scientist or technologist, uses well tested scientific and technological hypotheses, data and procedures, whereas a scientist cannot help hazarding speculations, and a technologist cannot avoid making recommendations on the basis of incomplete knowledge topped with intuition and judgement. The relevant distinction is not that between *episteme* and *doxa* but that between knowledge that is, or is not, a result of scientific (or technological or humanistic) research, as well as a possible subject of further investigation. In other words, the important distinction is that between research fields and belief systems. We shall return to this point in Section 4.1.

Another untenable dichotomy is that between *pure* and *practical* knowledge. Plato and Aristotle emphasized this distinction, praised the former and deprecated the latter. The pragmatists deny it, asserting that all "knowing is for the sake of doing" (Lewis, 1946, p. 3). However, some knowledge starts by being pure and remains so: it was acquired for its own sake and never finds application. Thus we do not "do" anything with cosmology, and nothing good has so far come out of the study of history—yet we cherish both and are proud of a culture that renders the cultivation of such disciplines possible and even desirable. However, it is hardly possible to forecast that a given piece of knowledge will remain pure forever. Electromagnetic fields, atoms, genes, fossils, brains, and many other things were initially studied with no practical purposes in mind, whereas nowadays entire technologies are concerned with using or manipulating them. In short, the pure-practical distinction does not amount to a dichotomy.

2.2. *Tacit and Explicit Knowledge*

Only some of our knowledge, whether sensory-motor, perceptual or conceptual, is conscious or explicit and expressible in some language: much of it is unconscious or tacit in the sense that we possess it without our knowing or without our being able to express it in any language. Thus craftsmen and artists, technologists and scientists, know how to do a number of "things" that they could not put into words. (The great Euler is reputed to have said that his pencil knew more than himself.) And any honest technology transfer deal stipulates that the customer will get not only artifacts and the corresponding operating and service manuals but also experts capable of teaching by example how to handle and service the system—precisely because not everything can be put into manuals.

A possible definition of the above concepts is this: If subject s knows p, then (*a*) s has *explicit* knowledge of p if and only if s also knows that s knows p or knows how to express p in some language; (*b*) otherwise s has *tacit* knowledge of p.

The difference has been explicitly drawn by a number of philosophers, in particular James (1890 I), and Russell (1918), who distinguished between *knowledge by acquaintance* (or *tacit* or *procedural* knowledge) and *knowledge by description* (or *explicit* or *propositional* knowledge). Polanyi (1958) emphasized the importance of tacit knowledge (which he called 'personal'); so did Oakeshott (1962). And both held that tacit knowledge is

superior to explicit knowledge. This is a curious dogmatic statement to make for, by definition, tacit knowledge is ineffable and therefore hardly subject to examination, in particular evaluation.

Both tacit and explicit knowledge can stay private or become public, depending on the knowing subject, the degree of advancement of the field of knowledge, and society. However, it is obvious that explicit knowledge, for being expressible rather than ineffable, is far easier to share—provided there are no social barriers to information. (Remember that nearly every society regards knowledge of certain kinds as esoteric, or accessible only to a small group of initiates, be they priests, politicians, or intelligence agents.) But all this is rather obvious and therefore philosophically unproblematic.

A problem of interest to the cognitive sciences and epistemology is whether there are neurophysiological conditions that impose the dissociation between knowing how and knowing that or whether, with some effort, we could render explicit every bit of tacit knowledge. The scientific investigation of this question has only started. We have learned that amnesic patients can learn certain skills even if they cannot remember the information gained by using them. Thus patients unable to remember day-to-day events (i.e. who have lost short term memory) can be taught mirror reading without their remembering which words they had read (Cohen and Squire, 1980). This finding suggests that the dissociation between tacit and explicit knowledge is quite deep-seated.

We still do not know, then, whether there must always remain a residue of ineffable knowledge. But at least we seem to have learned that the right way to approach this problem is through scientific investigation rather than by resorting to a ready-made doctrine (such as intuitionism, rationalism, or empiricism). The problem is not one of mere academic interest: it is also of interest to educationists and to experts in the transfer of technology. Indeed, both groups would like to know whether every bit of knowledge can eventually be poured into a manual or a tape.

The tacit (or procedural)/explicit (or propositional) distinction could be used to discuss the problem whether computers can be said to know anything. It might be argued that they can acquire know-how but not know-that: that their knowledge is purely procedural—e.g. that they know how to add integers but not what addition or integers are. However, if we stick by our psychobiological approach we cannot admit that computers know anything because, having no nervous systems, they cannot experience plastic changes in them. All we can say is that computers are in

some respects similar or analogous to knowing subjects, just as jets are in some respects similar to birds. These analogies are all-important for technology but of very little philosophical interest.

The tacit-explicit distinction intervenes also in current discussions in psycholinguistics. Empiricists claim that knowing a language is an instance of know-how in the same category as bicycling or hammering. Extreme rationalists, on the other hand, hold that knowing a language, like knowing anything else, is a case of know that (or propositional knowledge) in the same category as knowing that jets are fast or that the day has 24 hours. And Chomsky (1980) holds moreover that the basic grammar of a language is an inherited endowment somehow modified by experience and reason, perhaps in much the same way as songbirds improve by practice and imitation the songs they inherited genetically.

We will not repeat here what we held in Ch. 1, Section 3.1 about the inheritability of neurophysiological features that make it possible for a human to learn a language under suitable social circumstances. But we must pose the following problems. First, is it not likely that the young child's knowledge of language is a case of know-how in the same category as his knowing how to play with toys, and that such know-how turns into know-that only with formal education? Second, if knowing a language is to master the rules for generating grammatical sentences, and if such rules are innate or at any rate they are learned long before the speaker can form a notion of a rule, then are we really talking about *rules* proper, i.e. prescriptions for doing something, or rather about dispositions or even laws?

Whether the grammatical rules are inherited or learned as bits of know-how, they may be objective *patterns* ingrained in the brain rather than the constructs we learn at school. Thus knowing how to speak English amounts to having acquired the pattern of generating English sentences without necessarily *knowing* it, much as we learn to walk or play peekaboo. Likewise planets may be said to possess the laws of planetary movement without knowing them. Hence the expression 'knowledge of a language' is misleading if 'knowledge' is understood as being always conscious: it is mostly a matter of know-how rather than of know-that. Neither children nor illiterate adults nor the apes who have learned to manipulate human linguistic signals have explicit knowledge of language, hence they cannot make use of any grammatical rules in order to avoid or correct grammatical mistakes. They have only tacit knowledge of a language. Formal learning can bring this knowledge to the fore, i.e. transform tacit into explicit knowledge, patterns into rules proper.

Now two cautions against certain rather common confusions. The first is that tacit knowledge should not be equated with private or privileged knowledge (Section 2.1), for the former can often be shared or transmitted, if not verbally by example. And explicit knowledge may but need not be public: it may be kept secret. All public knowledge is or may be rendered explicit but the converse is false.

Our second caution is against mistaking public knowledge for objective knowledge, as explicitness and publicity are only necessary conditions for objectivity. Indeed we define objectivity as follows. First the general concept, applicable to both factual and formal knowledge. Let p be a piece of explicit knowledge. Then p is *objective* if and only if (a) p is public (intersubjective) in some society, and (b) p is testable (checkable) either conceptually or empirically. And now the strong or special concept of objectivity. Let p be a piece of objective knowledge. Then p is *factually objective* if and only if p has a factual reference (i.e. is about actual or possible facts).

Note again the difference between objectivity and intersubjectivity, a distinction that most epistemologists do not draw. (No subjectivist epistemology has any use for it.) Magical rules and religious dogmas may be intersubjective in a given society but they are not objective in the above methodological sense. As for the difference between objectivity *lato sensu* and *stricto sensu*, note that mathematical theorems are objective in the former sense but not in the second, whereas the statements made in science and technology are supposed to be objective in the narrow sense. (This point is of course being controverted in the social sciences, in particular in history. We shall take it up again in Vol. 6.) Finally, note that truth is not involved in objectivity. A statement may be objective and false, or true and non-objective. For example "$2 - 1 = 0$" is objective and false, and "I am not sure I'll ever finish reading this book" may be true but not objective. More on objectivity in Ch. 15, Section 2.2.

3. Belief

3.1. *Belief System*

In line with our biological approach to mind we construe beliefs, doubts and disbeliefs as brain processes of the mental kind. Like any other brain process of this kind, a belief has a genesis and it can change, even disappear, under the impact of other bodily (in particular mental) processes.

A possible characterization of belief and its cognates is as follows (modified from Bunge, 1980a, Postulate 7.3). Let b be an animal, u and v thoughts, and τ a time interval. Then:

(i) b *entertains* thought u during τ iff b has psychons that think up u, as well as other thoughts related to u, during τ;

(ii) b *believes* thought u during τ iff b entertains only u during τ (i.e. the psychons for other thoughts, related to u, are not activated during τ);

(iii) b is *doubtful* (or *undecided*) between thoughts u and v during τ iff b entertains both u and v alternately during τ, i.e. in such a way that the activation of the psychon for u inhibits that for v and conversely, i.e. cyclically.

Being brain processes, believing, doubting and disbelieving can in principle be investigated empirically (by neuropsychology). And, because the holding of certain beliefs controls social behavior, belief can also be investigated by social scientists. In fact it is being investigated by social psychologists and other workers. Such empirical approaches are of course at variance with the a priori approach adopted by the philosophers interested in the "logic" of belief (e.g. Hintikka, 1962).

Like any other brain processes, beliefs come in different strengths. We may grade them between -1 (maximal disbelief or rejection) and $+1$ (maximal belief or acceptance); and we may assign the value 0 to any thought that we are as strongly inclined to accept as to reject. More precisely, call B_b the collection of thoughts that animal b can initiate at time t, and T the set of all instants. We conjecture that there is some (presently unknown) function w_b from the cartesian product $B_b \times T$ into the unit real interval $[-1, 1]$, whose value $w_b(x, t)$, for x in B_b and t in T, is the *weight* or *strength* of x for b at time t.

Presumably, the values of w_b for different thoughts and times could be determined empirically. Such empirical research would show that belief strengths change in the course of time, occasionally in a dramatic fashion. If we can determine the collection B_b of thoughts of animal b, together with their respective weights, we have *eo ipso* determined the *belief system* of b, which we define as $\mathscr{B}_b = \langle B_b, w_b \rangle$.

To be sure, we have not determined the weight functions w_b but have merely expressed the hope that they may one day be determined empirically. As a matter of fact beliefs are already being subjected to experimental investigation, which is not surprising considering that they

are brain processes rather than inmates of an immaterial mind. Being brain processes they can be influenced by suitable cognitive stimuli as well as by the subject's own ongoing brain activity. For example, a subject led to believe that she had been touched with the leaves of a harmful plant (e.g. poison ivy), while in fact she had been touched with a harmless plant, may develop flushing, itching, and erythema (Barber, 1978). The findings of social psychologists, on the way belief can be manipulated, are, if anything, even more dramatic: see the classical studies of Sherif (1936) and Asch (1952).

Another problem concerning belief that has been investigated experimentally is the discrepancy between what we believe and what we *say* we believe. For example, in a given English speaking community most of the women may pronounce 'student' while claiming that they always say 'stjudent', which is the socially accepted pronunciation; men behave in the dual manner. (It is not that all those subjects lie, except to themselves: they may not be aware of the discrepancy.) In any case of this kind it is the experimenter's task to fill out the following matrix:

	Professing to believe or do p	Not professing to believe or do p
Actually believing or doing p	a	b
Actually not believing or doing p	c	d

with $a+b=1$ and $c+d=1$. If all subjects are perfectly truthful (to the experimenter and to themselves), $b=c=0$. In actual fact this situation is seldom encountered. I.e. most of us do not quite know what we believe or do—which is one more argument against the definability of knowledge in terms of belief.

To return to the belief weights w_b. Since beliefs are brain processes, in principle they might be assigned objective probabilities conditional on experience and circumstances. One might speak of individual b holding belief u at time t, with probability p (or of the individual's propensity or disposition to believe u at t). In such case we would set $w_b(u, t) = 2P_{bt}(u) - 1$.

And, of course, one might then introduce the conditional probability of a given individual switching from one belief to another. But, as noted, such probabilities must be objective, for it makes little if any sense to ask an experimental subject to reply to questions of the form 'What subjective (or personal) probability do you attribute to that proposition (e.g. the Schrödinger equation): 0.7, 0.8, 0.9, or 1.0?'. And they must be conditional: a person accustomed to evaluating beliefs critically is less gullible than one who has received a dogmatic education, and so the probability of his acquiring a belief in the ghostly is smaller than the average.

The preceding considerations do not carry over, without further ado, to groups of animals, such as professional groups or social classes. To be sure ideologists and even some social scientists sometimes claim that the bourgeoisie, or the proletariat, or some other social group, holds this or that belief, or has this or that belief system. From a psychobiological viewpoint this claim is false, because social groups do not have collective brains. (On the other hand the claim does make sense in the context of objective idealism, where there is a body of ideas floating around, which some individuals can grasp albeit only imperfectly.) However, the notion of group belief, or belief system of a group, can be rendered rigorous and useful in the following way. Form the intersection of all the beliefs of the individual members of a group, and the family of their belief strengths: i.e. collect all the beliefs that the group members share (hold in common), and assign each its weight. The result is this: The *belief system of group G* is the structure $\mathscr{B}_G = \langle \bigcap_{b \in G} B_b, \{w_b | b \in G\} \rangle$.

The intersection of the beliefs of the members of a group of animals of the same species is never empty. But if it is small, as it happens with every wide group and every intensely divided society, the concept of group belief is not of much use, and the social scientist is forced to subdivide the given group into more homogeneous groups. But when the intersection is large, as is the case with homogeneous groups, we can speak of *prevailing opinion*, or *climate of ideas*, or "spirit of the times" (*Zeitgeist*)—provided we do not forget that such prevailing opinion is neither held by a single individual nor by a ghostly collective mind, but is a construction out of individual belief systems.

We should not underrate the weight of prevailing opinion on inquiry. Indeed such prevailing opinion may be favorable, unfavorable, or indifferent to a given line of inquiry (or research project). If favorable, "it" will encourage the project to the point of overlooking blemishes and

discouraging competing projects. If unfavorable, "it" will exaggerate shortcomings and erect obstacles—which may either kill the project or force the investigators to be more ingenious, resourceful, careful, and critical. In either case, whether favorable or unfavorable, prevailing opinion is double-edged. (It works like a two loop control system, one of the elements providing negative, the other positive feedback.)

We are all sensitive, in some degree or other, to prevailing opinion and group pressure. Hence everyone is more or less suggestible and gullible, in addition to taking the good faith of others for granted. This makes us all more or less easy preys to opinion manipulation by teachers and preachers, politicians and ad-men. Contrary to widespread opinion, not even scientists are immune: the chemist who would not accept uncritically reports on success in synthesizing a certain molecule may fall in for homeopathy, psychoanalysis, or monetarism. A nationwide Gallup poll conducted in 1978 to assess the degree to which Americans believe in the paranormal and the ghostly gave the following results: 57% believe in UFOs, 54% in angels, 51% in telepathy, 39% in devils, 37% in precognition, 29% in astrology, 24% in clairvoyance, and 11% in ghosts. At about the same time it was found that belief in the paranormal is much higher among university students than among the public at large—a sad comment on the failure of the university to educate (Greenwell, 1980).

Why do so many people hold such groundless beliefs at a time when we are supposed to embrace science and technology? Presumably because some of those ideas seem to explain easily a great many phenomena, others are comforting, still others promise instant knowledge or power, and most of them have been received uncritically at an early age from the previous generation along with fairy tales and religious dogmas. In all such cases belief is supported not by argument or theory, let alone empirical evidence, but by tradition and a natural disposition to receive whatever one is given, as well as by the wish to secure cheap protection. (It is said that all risky and highly competitive professions are ridden with superstition. Sailors and airline pilots, truck drivers and stock brokers, soldiers and prostitutes often wear amulets.) The critical attitude and scientific inquiry are still recent and uncommon. Still, interestingly enough, few believers in those superstitions act consistently on the strength of such beliefs. For example, believers in astrology seldom choose their profession by consulting their horoscopes. There is always a gap between belief and action, principle and deed.

Experience is no deterrent to superstition because it can be interpreted in terms of the latter. On the contrary, we often acquire superstitious beliefs

from having experienced certain coincidences and from our unwillingness to admit chance (a very modern idea that has not yet seeped down). This has been suggested by a classical experiment on the origin of superstition in pigeons. Instead of reinforcing "good" behavior, Skinner arranged for food to be delivered to them at regular intervals regardless of the animal's behavior. (I.e. he gave them adventitious reinforcement, which is what we do with our children when feeding them cookies and TV when they return from school regardless of their grades.) It turned out that one bird was conditioned to turn about the cage, another thrust its head into the upper corners of the cage, a third developed a tossing response, and two birds a pendulum motion of the head and the body. "The conditioning process is usually obvious. The bird happens to be executing some response as the hopper [food dispenser] appears; as a result it tends to repeat this response" (Skinner, 1948, p. 168). True, there is some evidence that such superstitious behavior does not occur in rats unless they have hippocampal lesions (Davenport, 1979). Still, chance coincidence must play some role in the formation of groundless beliefs even among higher vertebrates. And among humans docile reception of traditional belief is often socially rewarded.

Uncertainty is unpleasant to most of us, and this is why dogma is so popular. Indeed the adoption of dogma removes uncertainty, for a dogma, or at least a clever one, has ready-made answers to all possible problems of a kind. On the other hand inquiry, particularly in science, technology, and the humanities, creates new uncertainties for every uncertainty it removes, as every research problem ends up, in the case of success, by posing new problems while solving some. Thus, inquiry does not decrease uncertainty. But of course this is not the unsettling and paralyzing uncertainty produced by the failure of dogma: it is the stimulating uncertainty that promotes further inquiry. (If only for this reason the information-theoretic model of the growth of knowledge is inadequate, for it equates such growth with information gain, which would in turn be the same as decrease in uncertainty.)

Having condemned dogmatism let us now sound a warning against radical or systematic skepticism. If we were skeptics all the time we would abstain from making choices and decisions, as well as from acting on them. Consequently we would soon die—ignorant to boot. Radical skepticism, in particular skepticism about all our own perceptions and thoughts, has no survival value. Here as elsewhere a middle course is advisable, namely *methodical skepticism*.

3.2. *Belief and Knowledge*

We saw in the preceding section that we do not always know *that* we believe or *what* we believe. And sometimes we profess to believe "things" we really do not know anything about, such as life after death, perfect rationality, and spotless democracy. Some people go even further: thus mystics and psychoanalysts claim that their doctrines must be believed in order to be understood. (Remember Isaiah's *Credo ut intelligam*—I believe in order to understand—and the psychoanalytic requirement that every practitioner be subject to analysis.) A few have gone to the extreme *Credo quia absurdum* (I believe because it is absurd). In all these cases belief is recommended as prior to understanding and testing. Believers in the supernatural or the paranormal are faithful to this maxim: rather than believing science they prefer to believe what flies in the face of evidence and reason. Thus Kierkegaard, realizing that Christian dogma is absurd, stuck to it and declared existence itself to be absurd.

The uneducated believes what he knows and even what he does not: he is basically a dogmatist, only occasionally an inquirer. The educated knows what he believes and does not quite believe what he knows: he is basically a skeptic, though seldom a systematic or radical one. In short, belief is not a condition for knowledge. On the contrary, knowledge is a condition for rational or justified belief (or disbelief or suspension of belief). *Knowledge, then, is not a kind of belief.* Yet, Russell (1948) and many other epistemologists have defined knowledge as a special kind of belief, namely as *justified true belief.* (More explicitly: s knows that p if and only if s believes p, p is true, and s is completely justified in believing p. See Lehrer (1974) for a fuller statement of the belief analysis of knowledge.) This view is refuted by all the cases of people who profess to believe what they really do not know, and disbelieve a number of "things" they know. A more sophisticated counter-example is this: The disjunction of an arbitrary proposition with a tautology is true, so according to the belief theorists we should be justified in believing it. However, such a disjunction cannot pass for knowledge, particularly if the nontautological component is arbitrary or indeterminate. Besides, according to the definition of knowledge in terms of belief, most scientific statements would not qualify as bits of knowledge because they are at best partially true (see Ch. 12, Section 1.1). Nor would abstract theories, such as set theory and group theory, qualify, for they are neither true nor false. Being inapplicable to the most interesting bits of knowledge, the belief analysis of knowledge is itself unbelievable.

A rational person will make some effort to know a proposition before believing or disbelieving it: his beliefs are special cases of his knowledge not the other way round. Consequently instead of defining knowledge as justified belief, a rational person will define justified belief in terms of knowledge. He may accept the following definitions:

(i) *s believes* $p =_{df}$ s knows p & s gives assent to p;

(ii) *s is justified in believing* $p =_{df}$ s knows p, and p is approximately true or has a model;

(iii) *s is justified in doubting* $p =_{df}$ s knows p and s knows of no grounds to assign p a definite truth value or a model;

(iv) *s is justified in disbelieving* $p =_{df}$ s knows p and s knows that p has been refuted or has no models.

The problem of the relation between knowledge and belief can be approached either descriptively or normatively. The descriptive study of cognitive attitudes, such as belief and disbelief, certainty and doubt, is a legitimate field of inquiry if conducted scientifically, as part of psychology. It is illegitimate otherwise, i.e. if conducted without regard for empirical testing. A normative investigation of cognitive attitudes, i.e. one serving as a basis for a rational theory of decision and action, should be empirical as well: it should be possible to check whether the norms it proposes are efficient, i.e. conducive to successful action. Clearly, any such normative theory must presuppose some notion of knowledge and some notion of truth, for othewise it may consecrate groundless belief, i.e. dogmatism.

One may go farther and demand that such a normative theory of belief contain the axiom that credences, or degrees of belief, be numerically equal to the corresponding degrees of truth. (That factual propositions are true or false to some degree or other is known to all scientists and technologists. For the notion of partial truth see Ch. 12, Section 1.1, and Appendix 3.) More precisely, we postulate that, if p and q are distinct propositions, and the degree of truth of p given q (as a baseline) is v, then the rational credence of p given (assuming) q should equal v. (On the other hand if p and q are the same, i.e. p is its own ground, and thus groundless except possibly for its consequences, then p should be assigned no credence at all.) A consequence of this postulate is that the truest propositions are assigned the greatest rational credences. Another is that, if p is true, then not-p is incredible. A third is that all the consequences of a true proposition are maximally credible. Clearly, any theory containing such axiom is normative not descriptive: it is a theory of rational or grounded belief.

The standard theories of knowledge and belief, namely epistemic logic and doxastic logic, first proposed by Hintikka (1962) and now active subjects of research (cf. Lenzen, 1980), are of a totally different kind. In fact they are a priori (i.e. "logics") and they do not use the concept of partial factual truth, so they are hardly relevant to scientific, technological or humanistic knowledge. To begin with, the basic concepts of epistemic logic, namely those of belief, certainty, and knowledge, are taken from ordinary knowledge. Surely they are symbolized, but symbolization is no guarantee of truth or depth. For example, the principle "If s believes p, then possibly p" remains hazy as long as neither belief nor possibility are adequately elucidated. Hintikka (1969) does elucidate the notion of possibility in model-theoretic terms: the outcome is the so-called possible worlds analysis of knowledge. However, the possibilities concerned are conceptual not factual, whereas in the case of factual knowledge we are interested in real (nomological) possibility. (See Vol. 3, Ch. 4, Section 2.) Secondly, epistemic logic is not interested in the grounds for holding beliefs—such as empirical evidence and harmony with previously confirmed hypotheses—and so it makes no contact with factual truth. Consequently it tolerates arbitrariness and irrationality, and so it is not a tool for examining beliefs with a view to justifying or criticizing them. Thirdly, epistemic logic deals with isolated propositions and arbitrary sets of such, which is not the usual case in science, technology, or the humanities. In sum, epistemic logic is not an efficient tool for analyzing knowledge, and thus belongs neither in descriptive nor in normative epistemology. It is just a *jeu d'esprit*.

It does not follow that Popper is justified in holding that belief and the study of it are unimportant. On the contrary, we live and die by our instincts and beliefs. In particular all our inquiries, valuations and conscious actions are guided or misguided by beliefs. Thus, we believe in the value of scientific research to know the world. Only, knowledge is not a special kind of belief, and therefore epistemology is not included in the study of belief; moreover, the latter belongs in psychology.

4. Inquiry

4.1. *Field of Knowledge*

Knowledge comes in degrees, and so does its dual, i.e. ignorance. Unlike knowledge, total ignorance cannot be assigned a brain activity of a definite type: it is the absence of such activity. On the other hand realizing, like Socrates, that one ignores something, is a bit of knowledge and therefore a

brain activity. There are two ways of overcoming ignorance. One, cheap but illusory, is to embrace uncritically some belief; the other, onerous but effective, is to engage in inquiry in order to obtain some first hand or second hand knowledge.

Consider the main steps in a typical process of inquiry.

Step	*Events in inquirer's brain*
1	Considers problem p.
2	Inquires about theoretical or empirical setting of p.
3	Thinks up possible solution s to p.
4	Makes preliminary check of s
5	Concludes that s is inadequate.
6	Searches for alternative solutions to p.
7	Comes up with candidate t.
8	Makes preliminary check of t.
9	Concludes that t may be adequate.
10	Searches for counterexamples to t.
11	Having found none (or having explained them away) declares t to be a plausible solution to p.
12	Decides to investigate t methodically.
13	Formulates tactical plan for investigating t.
14	Writes up proposal for research grant.
15	In the process finds new data or reasons relevant to problem p.
16	New information suggests reformulating or even radically changing p.
17	Above change forces modification of original research plan.
18	Writes up final proposal.
19	Works on problem while proposal is being considered.
20	Takes cognizance of rejection of research project on the ground that problem is worthless or not original, or that proposed method or suggested solution is inadequate or too original.

Inquiry is a particular kind of cognitive process: it is *directed cognition.* In fact inquiry stars with some problem found in a given fund of knowledge, it employs definite means, and aims at finding some thing or idea that may solve the problem. Inquiry takes then at least five items: a group of inquirers, an incomplete fund of knowledge, a set of problems (gaps in that fund), a set of research tools (conceptual or material), and a set of goals (cognitive or practical). Successful inquiry builds up an epistemic field as it proceeds: logic or geometry, physics or biology, engineering or

computer science, and so on. Since we shall make much of the concept of an epistemic field in this volume and the next, we may as well proceed to give a definition of the concept. (See Törnebohm (1979) for a different though congenial concept.)

The ten-tuple $\mathscr{E} = \langle C, S, D, G, F, B, P, K, A, M \rangle$ is an *epistemic field* (or *field of knowledge*) at a given time iff, at that time,

(i) C *is a system* composed of persons who have received a special training, hold strong information relations amongst them, and initiate or continue a tradition of belief or inquiry;

(ii) S *is a society* capable of supporting and encouraging, or at least tolerating, C;

(iii) the *domain* or *universe of discourse* D of $\mathscr{E}$ is the collection of objects of the inquiry, i.e. the reference class of the specific concepts of $\mathscr{E}$;

(iv) the *general outlook* or *philosophical background* G of $\mathscr{E}$ is composed of (a) an *ontology* or view on the nature of things, (b) an *epistemology* or view on the nature of knowledge, and (c) a *morality* concerning the proper ways of acquiring, diffusing and utilizing knowledge;

(v) the *formal background* F of $\mathscr{E}$ is a collection of logical or mathematical theories taken for granted in the course of the inquiry;

(vi) the *specific background* B of $\mathscr{E}$ is a collection of propositions and procedures, other than F, drawn from other epistemic fields;

(vii) the *problematics* P of $\mathscr{E}$ consists of problems concerning the nature, value or use of the members of D as well as problems regarding other components of $\mathscr{E}$;

(viii) the *fund of knowledge* K of $\mathscr{E}$ is a collection of propositions and procedures obtained by members of C at previous times;

(ix) the *aims* or *goals* A of $\mathscr{E}$ are of course the target (cognitive, practical or moral) of the specific activity of the C's;

(x) the *methodics* M of $\mathscr{E}$ consists of all the general and special methods utilizable in $\mathscr{E}$.

The first three components of this ten-tuple will be said to compose the *material framework* of the epistemic field in question. (Clearly this is a misnomer in the case of formal science, the humanities and theology, all of which deal primarily with conceptual objects not material ones. However, the expression has the virtue of reminding us that at least the first two components of every epistemic field *are* material systems.) The last seven components of the septuple shall be called the *conceptual framework* of the epistemic field under consideration. We shall return to this distinction in Ch. 14, Section 3 when dealing with paradigms and revolutions.

We distinguish two main types of epistemic field: *belief systems* and *fields of inquiry*. In a belief system the components 3rd to 10th hardly evolve or, if they change, they do so exclusively as a result of controversy or of external pressure: witness the ideologies. On the other hand in a field of inquiry every component from the 3rd to the 10th changes in the course of time, and does so mainly as a result of inquiry within the field or within contiguous fields. Besides, every field of belief is self-contained or nearly so: it is neither supported by, nor supports, other epistemic fields: witness pseudoscience. On the other hand every field of inquiry overlaps partially with some other epistemic fields. More on this in Ch. 14.

Belief systems do not always change spontaneously and gradually into research fields. On the other hand some fields of inquiry may gradually degenerate into fields of belief, and this in either of two ways. One is that the given epistemic field was finite to begin with: it had a finite number of problems, all of them soluble in a definitive way, and such that their solution did not pose any further problems. (We call *finite* any such field of inquiry even though we are at a loss to find clear instances.) The other way is by the gradual decay of inquisitiveness among the field workers, i.e. as dogma replaces conjecture and complacent belief takes the place of risky inquiry.

We do not know whether there are any finite and therefore exhaustible fields of belief: all of the known ones seem to be potentially infinite even when they attack finite problems such as censing populations or mapping chromosomes. Indeed, nontrivial knowledge of fact seems always to be spotty and more or less inaccurate—as well as obsolescent because of the extinction of things of some species and the emergence of things of new kinds. No sooner a body of data or of theory is pronounced perfect, than some gap or flaw is discovered in it—and somebody suggests ways of correcting or completing it. Therefore it may be assumed, though perhaps not proved, that, as long as there are workers left in a field of inquiry, the latter will continue to be cultivated—perhaps so intensively and effectively that it eventually turns into a different field. The antecedent of this conditional statement used to be taken for granted since the 17th century, but must now be questioned in view of the increasing virulence of antiscientific ideologies. More on this in Ch. 13, Section 4.

Strangely enough, though inquiry occupies an increasing number of people, and even though we are becoming increasingly familiar with it, there is hardly consensus on its very nature. To some, scientific, technological and humanistic research are arts like poetry or music, for

they demand imagination, dedication, and enthusiasm. To others, they are crafts that can be learned and practised by almost anybody. To still others, research is an industry—the knowledge industry. Finally, to many, scientific research and technological development are an immoral racket bent on destroying life. Curiously enough, there is a grain of truth in every one of these opinions. In fact at their best science, technology and the humanities are arts demanding very special callings, talents, and a passionate commitment. At their next best they are nothing but sophisticated crafts, only trickier than the craft of versifying or that of mixing sounds, but still crafts rather than arts. (Still, one draws the difference between mere craftsmanship, however superb, and the inspiration that alone can deliver original art, science, technology, or the humanities.) And it is also true that some research deserves being called an *industry*, namely when it deals competently with pseudoproblems or miniproblems. Finally, it is also true that some scientists, technologists and humanists are crooks—but they are only a few and they are easily found out. In sum, we do not know much about inquiry, beyond that it is interesting and many-faceted, hence worthy of being inquired into.

4.2 *How is Knowledge Possible?*

Is knowledge possible? If so, to what extent and how? The first question may sound facetious but it is a serious epistemological question and one that was not easy to answer before modern science and technology were born, i.e. at a time when most knowledge was mere opinion. Not only skeptics like Sextus Empiricus, Francisco Sánchez, and Unger (1975), but also dogmatists such as Plato and Nietzsche, have denied that we can get to know anything about factual matters. (Mathematics has seldom been questioned.) Their favorite gambit was to exhibit gaps and errors in the bodies of received opinion.

The classical mode of refutation of such radical skepticism concerning the possibility of factual knowledge is still valid. We show knowledge to be possible by exhibiting items of actual knowledge, thus making use of Aristotle's modal principle "If p, then p is possible". (*Trivial examples*: I know the address of my home, the tastes of my children, Maxwell's equations, the empirical formula of water, and the specific function of the heart.) This disposes of radical or *systematic skepticism*. And a consideration of its historical role shows it to be ambivalent for, although it can destroy dogma, it also discourages inquiry.

Methodical skepticism, or *fallibilism*, is a different matter altogether. It is the thesis that, although we can get to know something, such knowledge may be imperfect (neither complete nor entirely true), and is therefore susceptible to criticism and improvement. Fallibilism is a component of *critical realism*, i.e. the epistemology according to which we can represent and understand reality, though not exactly as it is. (More on critical realism in Ch. 15, Section 2.)

Fallibilism is not enough: it must be supplemented with another sobering element, namely a dose of *agnosticism*, i.e. the thesis that there *are* unknowns and unknowables. *Examples*: If the speed of light is indeed the ultimate speed, then we shall never know what is happening at the same time in a distant place; if the big bang hypothesis is true, then we shall never know what went on in the universe before that hypothesized initial explosion occurred; if the brain does not contain detectors and decoders of extremely low energy electromagnetic waves, then we shall never be able to communicate telepathically even at a short distance; if there are no laws of history then we shall never be able to forecast with precision any major social events.

Note the conditional form of our wagers: *If A (is true), then we won't be able to know or do B.* Such conditionalization is a protection against dogmatism. We do not wish to repeat Comte's and Spencer's mistake of drawing long lists of *ignorabilia* that turned out to be subjects of inquiry and even entire scientific fields, such as atomic physics and astrophysics. Comte and Spencer were merely projecting their own philosophical prejudices into the future. Any list of *ignorabilia* that we may care to compile must be based on current knowledge not current ignorance. Thus because we *know* that certain biospecies and human cultures have become extinct, we shall never get to know them adequately. (However, living fossils or remarkably well preserved remains may turn up if we do not give up the search.) Likewise because we *know* that every experiment is subject to accidental errors, we shall never know whether a particular measurement outcome is error-free. (However, we also know that such errors can sometimes be reduced by operating at extremely low temperatures, so there is hope.) In sum, our knowledge is and will always remain incomplete and inaccurate—but there is no knowing beforehand and exactly *what* will remain unknown. We do not know any laws concerning the evolution of knowledge enabling us to make such predictions.

We come now to the third and last of the initial questions, namely 'How is knowledge possible?' This is a classical problem in epistemology and

psychology, though one that skeptics (for obvious reasons) and others have refused to formulate. The classical answers to this question are:

(i) *naive realism*: the mind has direct access to things without the intermediary of either the senses or reason—e.g. intuitively or because the brain reflects the world;

(ii) *empiricism*: the mind receives impressions from the senses, so that it knows (only) phenomena (radical empiricism); the mind knows phenomena through perception, and it constructs abstract ideas (logical empiricism);

(iii) *idealism*: the mind grasps self-existing ideas (Plato) or essences (Husserl); there is a pre-established harmony between mind and world (Leibniz); the mind constructs the object, and certain ideas are prerequisites for any experience (Kant).

Although there is something to be said for each of these schools, none of them is fully satisfactory. Naive realism is too optimistic, it has no use for transempirical concepts and mathematical theories, it fails to explain creativity as well as error, and it proposes no cognitive mechanism. Empiricism underrates reason and is therefore impotent to explain abstract ideas such as those of logic, set theory, and abstract algebra. Idealism downgrades experience and does not account for error, let alone for the need to test our hypotheses. Besides, idealism and empiricism ignore the brain and the social context of cognition. Because they place mind out of both nature and society, empiricism and idealism can give no correct explanation of the generation of knowledge. Such reification of mind and its detachment from the world renders knowledge mysterious: in particular it makes science look "miraculously improbable" (Popper, 1972, p. 28). More on all these *isms* in Ch. 15, Section 2, where scientific realism will be defended.

Neuropsychology and the epistemology attached to it approach the problem in a radically different manner. We regard the organ of knowledge (the central nervous system) as a material system placed right in the middle of the natural and social world. As part of the world, the central nervous system can obtain some "inside knowledge" of it. The brain can represent other concrete items by being acted upon by them and generating neuronal configurations similar in some respects to the represented things. In short the brain can know matter because it is material itself. (The mythical immaterial mind, standing off the world, cannot interact with the latter, so it must be assumed to create the world all by itself.) Moreover, because cognition is a process in a concrete thing, it can study itself like any other

concrete process. And, because cognition is a biosocial process, it develops and evolves along with other processes occurring in the higher social vertebrates. Thus we can get to know something about knowledge with the help of biology, psychology, social science, and philosophy: the task of epistemology becomes feasible. But at the same time epistemologists must work much harder than heretofore: a priori pronouncements, even if insightful, won't satisfy us any longer.

The conditions that render knowledge possible are exactly those that have resulted in the emergence of animals endowed with a central nervous system. They are extremely special and therefore exceptional material conditions. If the universe were a high density solid block, no knowledge would be possible, not only because life would be impossible, but also because everything would hang together so tightly that knowing a little bit would require knowing the whole. Likewise, if matter were distributed much more thinly, the bonds among things would be so weak that no cells would self-assemble; and, even if they could emerge, they would be nearly isolated from the rest and thus would have no chance to communicate. Also, too short a life span would leave no room for learning: all perception and behavior would be controlled genetically. Besides such physical and biological conditions for learning, think of the social ones: the smartest animals are gregarious. Solitary animals can hardly learn from others. As a Leibnizian would say, ours is the very best world to be known—which does not entail that it was designed so that it could be known.

Finally, how about computers: do they know, and if so, is there any limit to what they can get to know? Most artificial intelligence practitioners claim that computers do know and that in principle there is no limit to what they can know. On the other hand the critics point out a number of limitations—e.g. that a computer can never know any statements lying beyond formal systems (because a computer program is a formal system describable as a Turing machine). Presumably, future computers could overcome or circumvent any such limitations *if* indeed they were capable of knowing *anything*. But they do not know anything if one admits any psychobiological notion of knowledge. The only way out for the computer cultist is to offer a redefinition of knowledge applicable to computers as well as to animals, and one not hinging on the mythical separation between body (or hardware) and mind (or software). So far such redefinition is not forthcoming. So far computers are only cognitive auxiliaries, albeit marvellous ones. They do not initiate inquiry and they do not create any new concepts, let alone hypotheses, theories, methods, or value systems. To

hold that computers do know is to indulge in metaphor—like saying that the emergence of every new "generation" of computers constitutes a quantum jump or a mutation.

5. Concluding remarks

Most of us go about inquiring in an empirical way, guided or misguided by intuition and tradition rather than by any explicit epistemological principles. Nonetheless we often manage to join the negative aspect of rationalism (namely its a priorism) with the negative one of empiricism (mainly the collecting of data for their own sake). Little wonder then that most of us harbor heaps of groundless beliefs and act accordingly, i.e. irrationally.

For example, we sometimes believe in the causal or even magical efficacy of the word even in the absence of interlocutors; we seldom reckon with chance coincidences and try to give causal explanations of them; we are prone to believe what we want to believe, and so we tend to disregard or reinterpret unfavorable evidence; we tend to fall for any comprehensive but simplistic and untested doctrine; and we have an excellent memory of our hits but a poor one of our misses. In short, we are often the victims of epistemic accidents.

It may well be impossible to avoid all the epistemic hazards. But with some formal education, some research experience and, above all, with some critical reflection upon our own successes and failures, we are bound to succeed in bringing down the frequency of epistemic accidents. And if we realize also that nearly every epistemology is but a blow-up of some partial truth, we may also avoid espousing it dogmatically, and may continue instead our search for, or updating of, an epistemology free from major blunders and capable of inspiring research instead of misguiding it.

CHAPTER 3

COMMUNICATION

Learning is private but inquiry is social, for it is done in some society or other. The social character of inquiry shows in our dependence upon tradition, in our need to share our questions and answers with others, and in our doing so through language and other, more primitive, modes of communication, such as gestures and calls.

Communication is an essential ingredient of social behavior in all gregarious animals. Through it we share cognitive attitudes and findings, and we influence cognitive processes in, as well as the behavior of, other animals. Through communication we are able to find out part of what goes on in the brains of our conspecifics and, what is more, we can modify such processes and therefore their behavioral outcomes. In this way every cognitive process is a component of some communication network or other—a network that nowadays embraces the whole of mankind. For this reason the lone knowing subject of traditional epistemology is as useless a fiction as the Robinson Crusoe of neoclassical economics.

Stressing the social character of inquiry and communication does not entail ignoring the fact that language is also a powerful thinking tool. As recognized long ago, thinking is to a large extent accompanied by talking to oneself. So, language discharges two functions: a cognitive and a social one. Chomsky (1975) and his followers have stressed the former, studying language exclusively as a mental faculty. On the other hand Wittgenstein (1953) and his followers are mainly interested in language as a mode of social behavior. Actually these points of view are mutually complementary rather than exclusive: speech is a tool of inquiry and learning as well as a mode of behavior and a social glue.

1. Subject and World

1.1. *Subject–Object Relation*

The pivotal notion of traditional epistemology is the relation between knower (subject) and known (object)—and, in particular, the relation between inquirer and world. Therefore a task of epistemology has always

been that of investigating the general problems involving such relation. To be sure, subjective idealism does not face this problem for it conflates subject with object, and so transforms the subject-object relation into a relation between the subject and himself. But in any non-subjectivistic epistemology the distinction between knowing subject and object known (or rather inquired into) is axiomatic. We call it the epistemological *duality postulate*—not to be mistaken for mind-body dualism. (For an examination of realism see Ch. 15, Section 2.2 and Lovejoy (1929).)

The epistemological duality postulate is essential to science, technology, and the humanities. Thus astronomy is the study of celestial objects not of the astronomer's inner life; sociology studies social systems, not the scientist's own "perception" of social facts; technology designs artifacts (physical, chemical, biological or social) instead of studying the daydreams of inventors; and the humanities study the lives and works of thinkers, artists, and men of action, not the scholar's own ideation or behavior. This does not entail that astronomers must ignore the way they perceive celestial objects, or that sociologists must push aside the way they conceive of social systems, and so on. Every specialist must study not only the referents of his field of knowledge but also the ways to study them properly, precisely in order to screen out the distortions ("artifacts") introduced by his viewpoints and techniques. Such self-knowledge, instead of the disregard for the knowing subject, will help attain or improve the objectivity that is supposed to characterize science, technology, and the humanities.

The distinction between subject and object need not, nay ought not to, imply their separation. I.e. the dual*ity* of inquirer and inquired need not imply a dual*ism* wherein the inquirer is divorced from his object of knowledge, and experience and theory are removed from nature and society, to constitute an independent realm. Such epistemological dualism, the partner of the metaphysical dualism of mind and body, is to be rejected if cognitive processes and their externalizations are to be understood as both organic and social processes, and therefore within the reach of scientific research.

A scientific approach to the subject-object relation is then at the same time epistemologically realist and ontologically naturalist (materialist). To be sure, historically not every realist epistemology has been associated with a materialist ontology. Thus Aquinas held consistently a realist epistemology and a hylomorphic and even supernaturalist philosophy of nature. Conversely, it is possible for a materialist not to hold a realist epistemology but to maintain, e.g. that all knowledge consists of con-

ventions and rules designed to facilitate biological and social adaptation. What does force one to adopt at the same time some form of epistemological realism and some form of naturalism is science not logic.

In other words, the subject-object distinction is correct and necessary as long as it is not transformed into belief in an utter separateness of man from his environment, as preached by the monotheistic religions and by idealism. Distinctive we are, separate not. In particular, science and technology are not detached from society but are social activities; and they do not stand across nature but are aspects of man's interaction with his environment. To put it in traditional jargon, the realist epistemology required by scientific, technological and humanistic research should be joined to a materialist ontology asserting the essential oneness, interconnectedness, mutability and lawfulness of all things, inquiring and inquired. (Note that we have written 'things' not 'objects', for there are non-things, such as constructs, that are objects of knowledge. More on this below.)

Another way of putting all this is as follows. The concepts of subject and object may be regarded either as mutually independent (not interdefinable) or as definable. Idealists have tried to define the object in terms of the subject (or of God). In particular, subjective idealists have attempted to define a material thing in terms of the sensations or the perceptions (or in general the ideas) it produces in a subject. Recall Stuart Mill's definition of a thing as a permanent possibility of sensation, and similar definitions by Avenarius, Mach, and at one time Russell and Carnap. Neither of them seems to have realized that any such definition is circular for including the very term to be defined. (Consider: "To be a *thing* is the set of all ideas elicited in act or in potency by the *thing* on some subject".) On the other hand objective idealism can be logically consistent but at the price of postulating a Supreme Being that exudes things as he thinks them up. By so doing the philosopher turns theologian and alienates himself from science and technology.

What about abstract ideas, such as those of triangularity and goodness? Do they not refute materialism and epistemological realism? They do so if construed as self-existing—i.e. provided one postulates immaterialism instead of trying to refute materialism by honest toil. The materialist has no trouble disposing of that opinion. He starts by noting that there is no evidence for the self-existence postulate: that all we know is that some animals are capable of thinking abstract ideas. And he goes on to postulate that such thoughts are brain processes, as assumed by physiological

psychology. So, when the subject thinks of triangularity he creates this idea (brain process) instead of apprehending it as in the case of perceiving a mountain. In other words, if the object is an idea then it is part of the subject. In this case strictly speaking—ontologically—there *is* no subject–object relation. In this case there is only a cognitive activity of the subject—in particular her feigning that the "objects" of her cogitations are self-existing.

This has important consequences. One is that not every object of knowledge is a material object existing independently of every possible knowing subject: some objects are parts (actual or potential) of some subjects. Another consequence is that, paradoxically, the knowledge of logic, mathematics or philosophy is, strictly speaking, a form of self-knowledge, because original logical, mathematical and philosophical objects are "produced" by the self instead of being self-existing. However, they have the appearance of objectivity and publicity because we have purposively stripped them of any idiosyncratic features. In other words, abstract ideas are formed by abstracting from personal characteristics and circumstances: they do not bear any birth marks and so are universal, i.e. thinkable in principle by anyone (Bunge, 1981a). In contrast, my knowledge of my present mood is only mine (unless I care to disclose it to somebody else) and, for being mine, it is not universal. Which brings us to the problem of the knowledge of mind.

There are multiple accesses to the brain processes of the mental kind, from physiological studies of brain activity and investigations of the behavioral manifestation of such activity, to external documents such as things fashioned or destroyed by the subject, to introspection or its verbal expression. The latter has been discredited by behaviorist psychology but it still is, and will always remain, an indispensable source of psychological knowledge, particularly when complemented and controlled by the other sources. Only, we must not say that 'the mind knows itself by introspection, i.e. by contemplating itself'. A linguistic behaviorist might say that this is bad grammar, because contemplation is an irreflexive relation: *a* can contemplate *b* only if *a* and *b* are different. We say that view is bad biology, for it reifies functions (activities). What can be said instead is that, in introspection, one part of the brain inquires into what another part of the same brain is doing, or what one subsystem of the brain did a moment ago: introspection, like self-consciousness, is just the self-monitoring of the brain. Moreover introspection, though necessary to understand the mind,

is insufficient, for it tells us only *that* we are having such and such subjective experience, not *how*, let alone *why*, we are having it.

1.2. *Inter-Subject Relation*

Because he can communicate in some form or other, every human being can teach others and learn from others—sometimes long departed ones, thanks to cultural externalizations such as written documents, drawings and tapes. This does not mean that cognitive attitudes, processes and their outcomes can be transferred the way goods change hands: learning something from someone else consists in going through certain similar brain processes that are not transferable and that most of the time cannot occur unless previous learning processes have taken place.

Communication is a kind of social behavior and it occurs in communication systems, be they relatively simple ones like the mother-infant system, or complex ones like an international television network. Every communication system is composed of three main components: sender, channel, and receiver. The system is active every time the sender emits a message that gets transmitted (usually with some distortion) through the channel, and the receiver succeeds in decoding it. Any of these components or even all three can be nonliving, but there is no communication system unless at least the sender is controlled, directly or indirectly, by some animal. (The receiver can be a robot. The sender too can be one, but it will function on behalf of some living being.)

Of all communication systems the most effective ones are those where transmitters are also receivers, i.e. where there is a two-way communication channel. The traditional classroom, with its lecturer and his passive pupils, was a one-way communication system. The modern class, where feedback is encouraged, is a two-way communication system that facilitates error correction as well as learning on the part of the teacher. Actually every society is, among other things, a communication system, for communication is a part of social behavior and therefore a member of the social structure (collection of social relations.) It is not only that cognition is influenced by social relations: every communication act, from asking a question to answering (or refusing to answer) it, from exchanging gossip to trading data, from listening to the news to selling a newspaper, is a social transaction that contributes to either social cohesion or social disintegration, to the growth or the decline of knowledge.

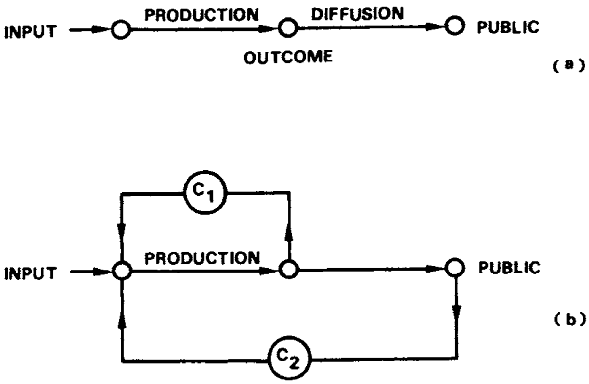


Fig. 3.1. Two views of the process of production and diffusion of knowledge. (*a*) Traditional views: open loop (no control). (*b*) Contemporary view: closed loops double control system: the subject corrects her own inputs (data, problems, etc.) in the light of her output, and society checks the whole.

From the point of view of systems theory, a communication system is a control system where the controls check the outcome of the process. Since there is the control by the knowing subject and by the society in which she is embedded, the system has two closed loops: see Figure 3.1. Without such control, irrelevance and error would go uncorrected, and society would neither encourage nor discourage any line of inquiry—for better or worse. Needless to say, the controls are in the hands (or rather brains) of all concerned, senders and receivers, producers and consumers of knowledge. This model contrasts with the open loop (no control) model of both empiricism and idealism: Table 3.1. The open loop model includes no mechanism, whether individual or social, for error correction, other than

TABLE 3.1
Three models of the production and diffusion of knowledge

	Empiricism	*Idealism*	*Realism*
Inputs	Data	Background knowledge	Background knowledge, data and problem
Transfer	Induction	Hypothesizing	Creative process
Output	Conclusions	New knowledge	New knowledge and problem
Control 1	—	—	Error correction
Control 2	—	—	Social checking

starting all over again, i.e. modifying the transfer function so as to obtain a different system. No such system change is necessary in the case of a closed loop control system.

We have avoided the term 'information' for several reasons. First, it is ambiguous, as it stands for both the act or process of conveying a message and for the content of the latter. Second, we communicate more than information or knowledge: we also transmit questions and commands, and we convey epistemic attitudes such as doubting, asserting, and denying. Third, since the inception of information theory (Shannon and Weaver, 1949), the term 'information' has acquired a technical signification that does not coincide with the ordinary language term, let alone with the traditional word 'knowledge'. Indeed information theory is concerned with the process of message transmission, not with what gets transmitted, which may be an equation or a nonsense string, a question or an order. The theory—like the communication systems it deals with, such as the mail and telephone services—is not concerned with the content of the messages the transmission of which it studies. Therefore information theory, though sometimes regarded as the basis of epistemology and semantics, is hardly relevant to them.

We shall have more to say on communication in Section 3 after having taken a look at the wider issue of the social matrix of the production and diffusion of knowledge.

2. Social matrix

2.1. *Cognitive Habitat and Niche*

Every animal has a habitat or region where it can live, as well as a niche in the ecological network. And every animal capable of learning has, within its total habitat, a *cognitive habitat* or region favorable to its cognitive activities. Thus at present the whole Earth and the Moon compose the human habitat, which may soon expand to encompass the entire solar system. But the Moon is not a good place to study anthropology, and Baffin island is not suitable for studying tropical zoology. Every human habitat is suitable, or unsuitable, to some degree or other, to learn about something.

In addition to a cognitive habitat, every animal capable of learning has a *cognitive niche* or set of possible cognitive states—or collection of "things" it can get to know. (A cognitive niche can be construed as a box in the corresponding cognitive space, or state space of the cognitive subsystem of

the nervous system. Every cognitive niche is included in the total niche of the animal.) Animals belonging to different species can have very different cognitive niches, which sometimes hardly overlap. Think of man, mole, and louse.

The cognitive habitat of an animal capable of learning includes its parents and all the conspecifics with which it is in touch. In the case of social animals such contacts are not sporadic and random but permanent and regular. In this case every animal can learn from every other member of its community, and can teach some other animals no sooner it has learned something as well as how to communicate it. Learning is a social enterprise. We do not start from scratch but from the findings (and limitations and failures) of our teachers, and we learn from one another. Tradition and the flow of information are particularly strong among humans, and they are getting ever stronger, particularly with the establishment of communication links and the invention of cognitive records such as drawings and writings. However, this social aspect of learning should not be exaggerated: we learn *in* society but it is the individual, not society, that does the learning, for society is brainless.

The union of all the individual cognitive niches may be called the *public cognitive space*, or the total cognitive space of society (or mankind). One advantage of an (ideal) free society would be that each member of it would have access to any region of this public cognitive space, i.e. to the cognitive abilities and achievements of his fellow men. This possibility does not exist in societies divided into groups (e.g. priestly, professional or military castes, political parties or corporations) that exert a monopoly on cognitive activities of certain types. Where the public cognitive space is not partitioned into water-tight cells, an individual can learn much and quickly.

Man does not evolve certain cognitive abilities except in certain societies. For example, hypothetical reasoning, or the ability to reason in terms of *ifs*, is the principal novelty in the cognitive development of the child reaching adolescence in urban communities. This stage is seldom reached by people living in the countryside or in backward villages, let alone in preliterate societies. Here the subject is incapable of adopting a detached attitude and of contemplating situations or views other than his own, or of taking an interest in questions that go far beyond his immediate field of interest (Piaget, 1972; Luria, 1976). Schooling, even in the village, has dramatic effects, not only on cognitive attitudes but also on affective ones such as punctuality, discipline, health, nutrition, family size, and sensitivity to

public issues—all of which have in turn a noticeable effect on economic growth (Lewis, 1955; Denison, 1962; Becker, 1964). In sum, human cognition is social through and through. Consequently psychology cannot study it properly unless it joins forces with sociology.

Take the social inhibitions, as well as the lack of social stimuli, to inquiries of certain types. A random sample must suffice. (*a*) The monopoly on literacy and inquiry, which was universal until recently, is a major obstacle to inquiry. (*b*) The extreme division of labor between manual and intellectual workers, and the low social status of the former, has been a major obstacle to scientific and technological innovation: craftsmen had no incentive to invent or to adopt new ways of increasing the quantity or the quality of the product, and intellectuals to engage in experimental inquiries. (In some nations, such as the U.S.A. and the U.K., scientists and technologists are not supposed to do their own machining.) (*c*) The prevailing value system may place little if any value on basic knowledge—this being why China, India and Islam did not contribute significantly to the building of modern science (Roche, 1976). (*d*) The prevailing ideology may discourage or even ban entire lines of research—e.g. of evolutionary biology in a theocratic state, and of social science in a totalitarian one. (*e*) A dogmatic school system will inhibit inquisitiveness and a critical attitude—whereas an excessively unstructured school will fail to teach intellectual discipline. (*f*) The preparation for, and waging of, war may accelerate the pace of destructive technology (drawing on a fund of basic knowledge). But it interrupts basic research by diverting human and economic resources—and by killing budding investigators in the battlefield. (*g*) The censorship exerted by organized dogmatism (church, party, clique, school) protects error from criticism and blocks the diffusion of new approaches and new results incompatible with dogma. Whatever does not fit with the latter remains unnoticed or, if noticed, is suitably "interpreted" (Fleck, 1935). (*h*) Intellectual fashion spurs research into the fad concerned but, by the same token, it diverts attention from other fields. (*i*) The current method of funding research encourages safe investigations expanding existing fields rather than opening up new ones. Indeed peer review tends to guarantee orthodoxy and suppress originality, encouraging researchers to swim in the main stream. (*j*) The requirement to remain within a given specialty makes for narrowness and blocks the diffusion of knowledge across cognitive fields. Trespassing should be rewarded not penalized, for it often gives rise to new approaches and occasionally even to new fields—as was the case with operations research, computer science, molecular

biology, and physiological psychology. (k) Society is not always quick to adopt improved ideas, procedures, or things. It will accept a novelty only if it is quite obvious that it fills a vacuum. Even so the resistance of people who have vested interests in the old must be overcome or bypassed: remember that "One person's innovation is ordinarily another's destruction" (Brewer, 1980). See Figure 3.2.

A few centuries of relentless growth of knowledge has made us forget that this is not a law but a recent trend, and that the right to learn, like any other right, is not a birthright but must be conquered or defended. In some societies most of the people are prevented from learning certain skills—if not by force at least because learning demands some leisure. Only a few societies recognize the right to know and make it effectively possible for everyone to learn. In most societies the individual is denied the right to learn about the major decisions that may affect him or her, and a fortiori to participate in the taking and implementing of such decisions. Every society has economic, political and cultural mechanisms for controlling the production and flow of knowledge: no modern society can afford to leave cognition totally in the hands of private initiative. (More in Ch. 13, Section 4.)

Having recognized the need to study the social matrix of cognition—and therefore the legitimacy of the sociology of knowledge—we should sound a warning against epistemic *sociologism*. This is the view that every cognitive activity and every bit of knowledge should be studied and evaluated *only* in

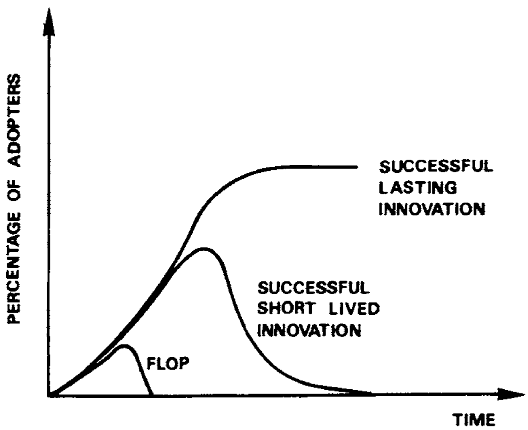


Fig. 3.2. The adoption of technological innovations.

the light of society and its history. Sociologism tells us that in the final analysis there is but one science: the science of man (Comte); that consciousness is determined (not just influenced) by socioeconomic conditions (Marx); that ideas are not so much in individual brains as in society—e.g. that Newton's physics was but a product of the first Industrial Revolution (Hessen, 1931); that the facts studied by science are not objective but are manufactured by the scientific community (Fleck, 1935); that truth is ultimately what the community of scientists decides to accept (Kuhn, 1962); that what distinguishes science from other cultural pursuits is its social condition (Ziman, 1979); and that an individual of genius is not one whose work is unique but, on the contrary, one "whose work in the end would be eventually rediscovered" (Merton, 1973)—so that a scientist of genius is "functionally equivalent" to a bunch of mediocre researchers. (By the way, Kuhn (1978) seems to have renounced sociologism to revert to internalism.)

Sociologism is wrong for the following reasons. First, all cognitive processes occur in some brain or other. Even a public utterance, or a formula written in a book, do not become cognitive items (i.e. processes in some plastic neural system) until somebody takes cognizance of it. Second, powerful new ideas and shattering discoveries are made only by exceptionally talented individuals: without these most of us would have no chance of participating, whether as beneficiaries or as victims, in any scientific or technological revolution. Third, although acceptance or rejection by a community are important to the fate of any new idea, popularity is not an indicator of truth anymore than impopularity an indicator of falsity: truth is a semantical and methodological category, not a sociological one. Fourth, although modern science and technology prosper only under exceptional social conditions, these come in a great variety, from capitalism to socialism, and from industrialization to semidevelopment. Moreover a mark of modern natural science is that its results hardly bear the seal of social circumstances, which on the other hand continue to exert deep influence on social science research. (For example, until the Great Depression economists were hardly concerned with unemployment, which since Keynes' *General Theory* (1936) has remained a crucial macroeconomic variable.) In short, society stimulates or inhibits inquiry in all fields, but does not do the inquiry because it has no brain. We do not have to choose between the isolated knowing subject of classical epistemology and psychology, and the all-powerful society imbued with a spirit of its own: we can choose to study the more modest but real inquiring subject-in-society. See Figure 3.3.

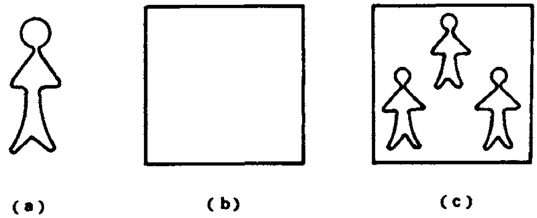


Fig. 3.3. Three views of cognition. (*a*) Traditional epistemology: the isolated knowing subject. (*b*) Sociologism: the brainless social framework. (*c*) Social neuropsychology: the community of learners embedded in a wider social matrix.

2.2. *Learning Community*

Every human being is a member of some *inquiring system* or *learning community*—i.e. a system of persons who learn by themselves as well as from one another by exploring and thinking, imitating and teaching, questioning and criticizing. And every learning community, whether advanced or primitive, has one or more *learning professionals*—medicine men or scribes, craftsmen or engineers, scientists or philosophers, teachers or journalists. Although these professionals may discharge other functions as well, their specific function is that of acquiring, utilizing or diffusing knowledge. These learning professionals form a subsystem of the cultural system that composes every community regardless of its level of development. Each of them holds information relations with other members of the community: some of them direct or interpersonal, others indirect, i.e. via couriers, correspondence, or publications. The potential clientele of these learning professionals is society at large.

In a modern society every learning community is divided into four main subsystems: the scientific, technological, humanistic, and artistic communities. Each learning professional is likely to be a member of several subsystems of the total learning community: e.g. a laboratory or some other unit, a university (or firm or government) department, a chapter or a professional society, an "invisible college" formed by knowledge consumers and evaluators, etc. In these conditions no learning professional is free from influences or completely devoid of cultural power over others. Typically, the crackpot is marginal: he is deprived of the benefits of stimulation and feedback. ("It is astonishing what foolish things one can temporarily believe if one thinks too long alone": Keynes, 1936, p. xxiii.)

The active member of a learning community shares in upkeeping certain inquiry traditions, updating them, and introducing new ones.

The organization of society as a whole, as well as that of each of its learning communities, is no less important for the conduct of inquiry than native intelligence, motivation, and industry. A mediocre intelligence can make substantial contributions to knowledge provided she is a component of a well organized learning community, whereas a genius is likely to go to waste in a cultural desert. The quick pace of innovation in science, technology, the humanities and the arts since the dawn of the modern era can be explained in terms of the emergence of a new type of society, more receptive to novelty and more mobile than the feudal society. The earlier explosion of knowledge that accompanied the birth of civilization about 5,000 years ago, and the no less dramatic explosion that accompanied the emergence of agriculture about 15,000 years ago, are further examples of the connection between modes of inquiry and social structure. Also, the revolution in technology and art that occurred about 30,000 years ago is likely to have accompanied a social reorganization which may have been prompted by a steep increase in population density, which in turn must have strenghtened and quickened the information flow.

Every specialized learning community, such as the community of biochemists or that of philosophers, is a social system. As such it can be represented by the triple *composition–environment–structure* (Vol. 4, Ch. 5). The *composition* of the community that cultivates a given field of science, technology, art, or the humanities, is of course the collection of all the workers in that field plus the support personnel (technicians, clerks, librarians, etc.) and their work tools. The *environment* of the community is the collection of all the items in the natural and social environment of it, including the subject matter in case this is material (e.g. the biosphere in the case of biologists, the whole universe in the case of cosmologists). And the *structure* of the community is the collection of relations of inquiry, teaching, receiving or conveying information, managing, etc., that give the learning community whatever cohesion and efficiency it may have. (We say that these activities are *relations* because this is what they are from a logical point of view. Thus inquiry is a ternary relation, for the standard form of a statement including the concept of inquiry is this: A inquires into B by means of C. Likewise teaching: A teaches B to do C.)

The study of learning communities is of course a task for social scientists, but their study as sociosystems has only begun. However, in order to understand any such community one must be able to understand what its

members do: what problems they face, how they handle them, what kinds of hypotheses, data and methods they use, and so on. Unfortunately all too often the anthropologist or sociologist studying learning communities ignores this caveat and limits his study to observing the motions through which his subjects go—which is compatible with behaviorism but far off the mark of a scientific study of cognitive activities. Not surprisingly, the results of such studies are trivial or even grotesque. Ignorance is no substitute for inquiring about inquiry.

Every normal social system is propelled by cooperation as well as competition. The learning communities are no exception. Cooperation includes in this case help in transmitting skills and information, formulating or reformulating problems, devising hypotheses or methods, offering constructive advice or criticism and, in general, sharing knowledge. Competition includes making destructive criticism, proposing rival theories, methods or data, and fighting over the support of third parties, such as assistants, colleagues, and funding agencies. Cooperation gives individual workers the necessary support, and competition keeps them on their toes; competition helps find out error, and cooperation corrects it. Where cooperation falters, the system disintegrates; where competition disappears, the learning system may degenerate into a belief system.

In an increasing number of fields inquiry is conducted by *research teams* animated by an *esprit de corps* and where the division of labor, sometimes excessive, is compensated for by efficient organization. The research team, particularly conspicuous in natural science and technology since World War I—but still exceptional in the humanities—was a natural offspring of the scientific enterprise led by an autocratic leader. The research team is often democratic and can handle not only cross-disciplinary research but also unidisciplinary inquiry involving a number of skills unlikely to be mastered by a single individual—e.g. theorizing, computing, designing experiments, measuring, and performing statistical analyses. The publication summarizing the results of the inquiry is usually signed by the principal investigators, who acknowledge the help of their assistants, or by all the members of the team even if they number twenty or thirty individuals. Team research is far more common in experimental and applied science than in theoretical science. It is most successful when research can be planned in detail—which is possible only when known methods are employed to attain specific goals in a planned manner, which is not the case of the exploration of the unknown (Abelson, 1965).

The *research factory* is a different kind of inquiring system. It is usually set up by a captain of industry or a government official who controls

resources, hires and fires, establishes the goals, and sketches the research or development project. Edison's Menlo Park, the Manhattan Project and the Apollo Mission are good examples of successful research factories. They were successful because their members agreed on a corpus of scientific or technological knowledge, on the proper methods to test new ideas and new artifacts, and on the ultimate goal. When such fundamental agreement is missing, mediocrity or paralysis is bound to set in. Although most research factories do not congregate large numbers of star researchers, they do ask for the advice of outside consultants capable of solving some of the toughest problems encountered by the team. Because research factories are mission-oriented, they discard any new problems and new ideas unlikely to be useful to attaining their goal. (Those who shun originality seldom get it.) Moreover the investigators employed by industry are not free to publish their findings, for the latter belong to their employers, who are anxious to patent and exploit discoveries, not to make them available to society. So, research factories curtail both the freedom of inquiry and the freedom of communication.

A third type of inquiring system is the *committee*. Committees may produce, analyze and evaluate research policies and projects, but of course they do not conduct original inquiries. Because most committees are watchdogs rather than explorers, they tend to be conservative. Galileo was condemned by a committee, and Einstein by political bureaus; likewise Marx and Darwin are still being condemned by churches and political groups. Since most people are afraid of novelty—either because they do not understand it or because they have vested interests in the old—the best result a committee can come up with when faced with radical novelty is an ineffectual proposal—such as referral to a subcommittee. Democracy has yet to solve the great problem of combining expert judgment with rule by committee. The key is agreement on both goals and means—a tacit agreement so much in evidence in the thinking and action that went into the selection of the horse and the camel. (It is false that a camel is a horse designed by a committee. What is true is that a committee of inexperts looks as though it were designed by a horse or a camel.)

The local, national, regional and international learning communities are the trustees of knowledge and they provide the mechanism for the steady production and diffusion of it. At the time of Galileo and Descartes the scientific community, then concentrated in half a dozen countries, counted about 200 scholars, mostly amateurs. (Armand Beaulieu, personal communication.) Today the scientific and technological community is spread over the entire world and is at least two million strong, not counting the

support personnel. These huge numbers facilitate cooperation and intensify competition. The outcome is not only greater productivity but also higher and higher empirical standards. The ultimate test of a scientific, technological, humanistic or artistic innovation is its achieving international recognition.

The existence of learning communities explains simultaneous discovery and invention, which are otherwise mysterious. Popular explanations for such events are that certain ideas are "in the air" (objective idealism), or that they are necessitated by society, particularly the economy (vulgar materialism). The truth is that coinventors or codiscoverers share (*a*) certain brain properties, (*b*) certain work habits and values, (*c*) a tradition, and (*d*) a set of problems. If half a dozen talented people share a body of knowledge and values, and tackle a given problem at about the same time, it is not unlikely that a couple of them will come up with the same solution at about the same time. Cultural parallelism is no more mysterious than genetic parallelism.

We wind up this section with a couple of remarks. The first concerns group learning: it is sometimes held that not only individuals but also human groups can learn (Botkin *et al.*, 1979.) Literally taken this is impossible, for groups do not have a brain of their own. What is true is that (*a*) individuals in a group can learn from one another, and they can learn to think, feel and act so as to optimize the efficiency and the cohesion of the group, as a result of which (*b*) the group evolves.

Our second remark concerns the ontological status of inquiring systems. At first sight they are not concrete or material entities because they lack most of the properties characterizing physical or chemical systems, and because they make use of constructs. However, (*a*) any system composed of material entities, such as human beings, can be said to be material itself, and (*b*) inquiring systems are not composed of persons and constructs on a par but only of persons endowed with brains capable of conceiving constructs. The material/conceptual dichotomy is methodological, not ontological: it is only a convenient fiction to abstract ideas from the brains capable of thinking them up. (Cf. Bunge, 1981a.)

3. Communication

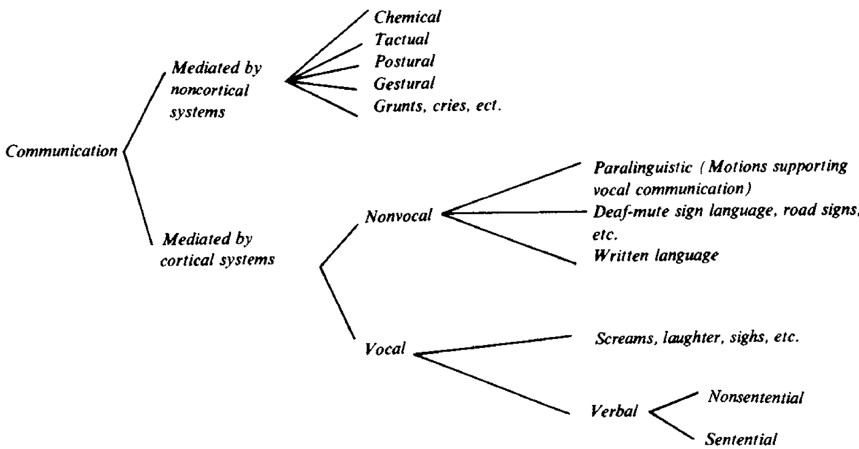
3.1. *Communication System*

All animals endowed with a central nervous system, from ant to primate, seem to be able to communicate, i.e. to send and receive signals carrying some information. (Cf. Sebeok, 1977.) Communication is particularly

developed in social animals, where it discharges at least the following functions (Lindauer, 1971): (*a*) to inform other animals that they belong to the same species or even the same society; (*b*) to inform others what their position in the social organization is; (*c*) to inform others in what state the individual is (e.g. whether it is ready to court, fight, or communicate where food can be found); and (*d*) to give or receive instructions concerning tasks of social utility (e.g. to bring in food or take out refuse). Communication, then, is a powerful tool of biological adaptation and social organization. However, here we are mainly interested in communication as a means to acquire, elaborate, and share knowledge.

The key concepts involved in any comprehensive discussion of communication are those of signal, message, significance, understanding, and communication. Definitions of these concepts in psychobiological terms are proposed elsewhere. (*Treatise*, Vol. 4, Ch. 4, Def. 4.55, and Bunge, 1980a, Ch. 9, Def. 9.8.) Here we shall concentrate on communication systems and, in particular, languages. A *communication system* is a concrete (material) system composed of animals of the same or different species, as well as nonliving things, in some environment (from beehive to the universe), and the structure of which includes signals of one or more kinds. Signals are processes executed or controlled by some animals and capable of carrying messages to other animals able to decipher them. These signals must be physical or chemical processes: the mythical psychical waves do not work because they do not exist. And the communicating animals must be endowed with suitable sensors: there is no extrasensory communication.

There are a great many kinds of signal employed in animal communication:



We are more interested in signals systems than in occasional stray signals such as grimaces, screams, or grunts. A *system of signals*, or *communication method*, is a (natural or artificial) collection of interrelated signals, involved in some communication systems, and allowing an animal to compose and convey messages addressed to other animals (primarily but not exclusively conspecific). A system of signals can be either inherited or learned, it may or may not contain a set of basic signals (e.g. words), and it may or may not be composed of an indefinite number of signals. Human speech and the American Sign Language, drawing and sculpting, pictographic and phonetic writing, musical notation and chemical notational conventions, mathematical and computer languages, are artificial systems unique in the animal kingdom.

Take for example the American Sign Language (ASL). Its expressive power is said to be as great as that of ordinary spoken language, though its rate of production is half the rate of the latter (Bellugi and Klima, 1979). Another example of a language is chemical nomenclature. The symbols for the 100 or so chemical elements, together with a set of numerals, suffice to form (by concatenation) the stoichiometric formula of any chemical compound. But in this case the syntax is neither conventional nor a product of a haphazard historical development: it summarizes chemical laws.

A *language* is a system of signals that are invented or learned rather than inherited. Moreover every signal in a language is either a simple signal or a complex one formed by concatenating two or more simple signals according to definite (syntactical) rules. There is no limit to the number of signals (hence messages) that can be formed in a language out of simple signals (e.g. words). (Mathematically speaking, a language is embedded in a free semigroup on a set of basic signals. Caution 1: The standard description of a language as a free semigroup is incorrect because in actual fact we can form only finite collections of strings of symbols of finite length. Caution 2: Some English words do not combine associatively. Thus '(small car) buyer' is not the same as 'small (car buyer)'.)

Many systems of signals, such as road signs, the human "body language", and the signals used by social bees, are not languages proper. Neither of them has a syntax, and the last two are inherited or instinctive rather than learned. Nevertheless some students of animal communication regard the systems of signals used by bees and ants as sets of conventions. Thus they state that taking the sun as a reference point, and "reading" the number of waggles or sound bursts as measuring the distance to the food source, are so many conventions. (See, e.g. Brines and Gould

1979.) But surely this is a misuse of the term 'convention', for it is admitted that such "rules" are wired in the insect's brain.

Human languages, on the other hand, are multiple, conventional, learned, and changeable. They are conventional because there is no law of nature commanding to name things in any one way, or placing noun and adjective in one order rather than another, and so on. (Caution: Although natural languages are conventional they are not the product of deliberate agreements but of cultural evolution.) Consequently any given linguistic convention can be replaced with another or even dropped altogether. And new conventions can be introduced provided they satisfy certain conditions ensuring over-all consistency as well as respecting what the French call *le génie de la langue*—whatever this may turn out to be.

So far as we know only humans have invented conventions and, in particular, linguistic conventions. However, there are indications that anthropoid apes can learn some of our linguistic conventions. A number of workers (Gardner and Gardner, 1969; Premack, 1971; Savage-Rumbaugh *et al.*, 1978; Patterson, 1978) contend that chimpanzees such as Sarah, Lana and Washoe, and gorillas such as Koko, have learned certain human languages such as ASL. They would master them to the point of communicating with their trainers as well as with some conspecifics, and to convey new thoughts not taught them, as well as to talk to themselves. The critics claim that these animals just respond to their teacher's previous utterances, that they have learned to use certain signals by sheer conditioning, and that the system of signals they use is devoid of syntax (Terrace *et al.*, 1979). The whole thing, according to them, would be a delusion, like the Clever Hans phenomenon, resulting from careless methods and anthropomorphic interpretation (Umiker-Sebeok and Sebeok, 1981).

The linguists who have become involved in the controversy—and there are not many—tend to disbelieve the thesis that beings other than man can make use of language. In particular, the transformational linguists pose the problem as follows. The fountain of a language is its grammar, which is composed of syntactical, semantical, and phonological rules for the production and understanding of correct sentences of the language. If an animal is capable of forming new correct sentences—new, that is, to the animal, and which is has not learned by rote—then it can be said to have mastered (to some degree) the grammar and thus the language in question. If on the other hand the animal does not form consistently any new and correct sentences, this is a clear indication that it does not possess a set of

rules (a grammar) to generate the sentences of the language. Now, the empirical evidence for "ape language" is weak: it refers to a handful of apes and there is no hard evidence that they do produce sentences really new to them, rather than repeating sentences formed by their trainers. Some workers hold a stronger view, namely that (*a*) there is no favorable evidence whatever, and (*b*) there *can* never be any such evidence, for language is built into the human mind—perhaps to the point of not being learnable at all. Needless to say, there is no evidence for these countertheses either.

It would seem that the investigation into "ape language" should start all over again and be conducted in a more rigorous and thorough manner. In particular the following points should be taken into account. First, a more adequate definition of a language (in particular a human language) should be proposed. (I do not mean a more exact definition, for the one offered by the mathematical linguists, namely in terms of the concept of a free semigroup, is exact. What I mean is a definition fitting any ordinary language.) Second, not only apes but also other animals, in particular "talking birds" and songbirds, should be investigated. Third, the performance of apes and other animals should be compared to that of children for, after all, we hardly know how a child acquires language (Bindra, 1981). It may well turn out that a child's speech, too, is cued by his interlocutor, who may in turn be a victim of the Clever Hans phenomenon. Fourth, the following hypothesis is worth being investigated. When an animal learns to name things or events, it may proceed as follows. In a first stage the object and the symbol (word, chip, figure, etc.) elicit each a process in some neural system. Learning to name the object by the symbol consists in connecting the two psychons or in establishing a third psychon for both. And this is noting but conditioning, that *bête noire* of mentalism: Figure 3.4. As for new sentences, they may result from the sequential activation of psychons for learned words, or from the spontaneous activity of psychons of new types. Since apes and other animals have given ample evidence of inventiveness in other fields, it should come as no surprise that they can learn sentences and even form new ones. But this hypothesis has yet to be confirmed by carefully controlled experiment.

Why has the problem of "ape language" stirred such spirited controversy? There are several reasons for this. One is that the problem is tough, the evidence inconclusive, and our knowledge of the way humans learn to speak is not far superior to our knowledge of the way Sarah, Washoe, Lana, Koko and a few other cousins have learned to communicate using man-made signals. Another reason is the novelty of the

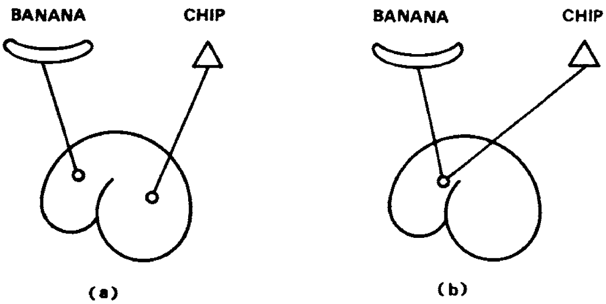


Fig. 3.4. Learning to name an object. In a first stage object and symbol activate different psychons. In a second stage a single psychon is activated.

thesis that animals other than men can talk, particularly in view of the traditional definition of man as "the talking animal". If true, the thesis would have profound implications for psychology (particularly psycholinguistics), anthropology (particularly palaeoanthropology and prehistory), and evolutionary biology (in particular that of primates). A third reason is that the thesis has ideological overtones. Indeed if the thesis is false, then it becomes easier to defend the view that the human mind has nothing to do with subhuman brains: there would be an unsurpassable chasm between man and beast, which evolutionary biology could not hope to explain. Thus the refutation of the hypothesis of "ape language" (or rather that apes can be taught some human languages and the mode of thought that goes with them) would bring comfort to the various religions and spiritualistic philosophies that hold the dogma of the spiritual (and possibly partly supernatural) nature of man. Actually if the "ape language" theses were refuted the materialist need not concede defeat: he could still hold that language and thought are brain functions that have evolved biologically and socially from humbler neurophysiological functions.

Unfortunately we know next to nothing about the origin of language and, because any prehistoric and palaeoanthropological data we may get hold of will at best serve as circumstantial evidence, we may never be able to formulate anything beyond clever conjectures. (Cf. Lieberman, 1975; Harnad *et al.*, 1976.) However, since language is *inter alia* a means of communication, it is reasonable to conjecture that every major leap in the evolution that led from our remote ancestors, a few million years ago, to *Homo sapiens sapiens*, was accompanied by a major advance in communication. It seems also reasonable to hypothesize that the language of

primitive man was quite different from ours: for one thing it may have been just a collection of words, and thus as agrammatical as baby talk. There is no reason to believe that it satisfied the definition of a language fashioned to fit modern languages. Unfortunately most linguists seem reluctant to accept any alternative definition, and thus they close the door to the very investigation of the evolution of language.

To return to the problem of language acquisition. Thanks to the work of developmental psychologists—Piaget in the first place—and neurolinguists, we are learning a bit about the development and loss of linguistic abilities. For example, it seems that language and substantive knowledge are functions of different brain systems. This is suggested by familiar experiences such as forgetting what a name stands for, and what the name of a known concept is. It is also suggested by the observation that a (semantic) aphasic knows that a cup is to drink from even if he is unable to name it. Moreover the surgeon can cut some of the neural links between speech and thought, as shown by commissurotomy patients, who are unable to say some of what they think. (A right-handed person who has had his corpus callosum cut cannot react verbally to verbal stimuli projected to his right hemisphere, but he may produce nonverbal responses indicating that he has understood the message conveyed by the stimulus.)

Furthermore it is possible that syntax and semantics are in charge of different psychons, as suggested by the finding that some neurological (and philosophical and sociological) patients produce gibberish sentences that are grammatically well formed. Also, reading and writing would seem to be controlled by different brain centers. Indeed severe dyslexia patients can write sentences which they are unable to read. Neurophysiological examination points to lesions in the angular gyrus (and usually also in the splenium) as causing dyslexia without dysgraphia. And lesions in the supramarginal gyrus have been shown to cause dyslexia with dysgraphia.

Further confirmation of the hypothesis that language and knowledge are functions of different brain systems is provided by the well known fact that illiterates and deaf-mutes can learn to play chess, paint, draw blueprints, and read maps, neither of which involves verbalized thinking. (Chess thinking is imagining moves, and paint thinking is imagining shapes and colors.) The very existence of nonverbal or nonsymbolic thought refutes the view, rather popular among philosophers from Plato to Quine, that all thought is symbolic or verbal, so that (methodological consequence) the study of language exhausts that of thought. Once more, empirical research has killed philosophical speculation.

Thought and language are then different. But they are coupled: learning to speak goes hand in hand with learning to perceive and to think, and each aids the other. Besides, using a language is a cognitive ability, so that knowing a language is a kind of knowledge, though not one of subject matter. Also, the more substantive knowledge we want to carry, the more linguistic container we need. In general all cognitive abilities—and language is one of them—are linked, and linguistic dexterity helps acquire and perfect other cognitive skills. For example, bilingual children all over the world do better at problem solving and "divergent thinking" (inventiveness) that unilingual ones (Lambert, 1981).

Our knowledge about language acquisition (and unlearning) is insufficient to give definite answers to certain questions that have given rise to a great deal of controversy in recent times. (The temperature of a debate is an ambiguous indicator: it may indicate profound ignorance of the subject, ideological implications, or both.) One such question is whether there is a special "capacity" for learning language—assuming we know what a "capacity" is. Whereas empiricists claim that language is learned like anything else, innatists hold that humans are born with a special capacity for learning languages—and some of them go as far as claiming that we are born knowing the main features of language. These conflicting theses are difficult to evaluate because of their imprecision: what does "general (or special) learning capacity" mean? The problem becomes clear when reformulated in neuropsychological terms, e.g. as "Are there any special neural systems ("areas", "structures") that do the language learning and no other?" The first conjunct of this problem has a definite answer in the affirmative: the Wernicke and Broca "areas" are the ones in charge of verbal language. (However, their location varies considerably with the individual.) But the second conjunct, namely the question whether such "areas" cannot learn anything extralinguistic, is yet to be answered or perhaps even asked.

A second controversy is closely related to the foregoing and it concerns the way we learn languages (cf. Sampson, 1978; Stich, 1979; Chomsky, 1980). Is it a step by step process, essentially one of trial and error operating on a blank slate (*tabula rasa*), as empiricism maintains? Or are we born with tacit knowledge of a Universal Grammar—common to all natural languages—as the innatists (or rationalists) hold? The latter have attacked linguistic empiricism holding that it does not explain why children learn so quickly the correct grammar: if they were to learn by trial and error they would never learn to speak at all, for it would take them an eternity to

discover all the rules of the language from any given set of linguistic data. True, but the falsity of empiricism does not prove rationalism—the more so since there is not the slightest evidence for the inheritability of anything beyond the mere capacity or potential to learn languages (or farming or astronomy). What we do know instead is that all normal humans share certain neurophysiological characteristics and are born into some culture or other; and that both the *neural universals* and the *cultural universals* determine the ways we learn. Some of these, which are properties or regularities of concrete systems (neural and social respectively), could be the linguistic universals that Chomsky and his followers attribute to an abstract (and yet to be discovered) Universal Grammar.

To conclude this section, I submit that the problem of language learning is a scientific not a philosophical one, and therefore neither of the two philosophies involved, namely empiricism and rationalism, is competent to solve it. We still know too little about the matter and won't learn more unless the acquisition of a first language be studied as a biosocial process, one of simultaneous neural development, socialization, and enculturation. (See Hebb *et. al.* (1971) for suggestive psychobiological hypotheses on first language learning.) We might have a satisfactory answer if the problem had been treated as one in biological and social psychology rather than in pure (nonbiological and nonsociological) psychology, let alone as a puzzle in the prescientific philosophy of mind. (See Bunge, 1983.)

3.2. *Language, Thought, and Reality*

Regardless of its unknown—though undoubtedly humble—origins, today's human language is unique. Unlike any other systems of signals, or communication methods, human language has become a *general purpose device* (James, 1890, II, p. 356). Indeed it can convey (if suitably enriched when needed) any message we may think of; it serves as a social link (to exchange information and modify other people's behavior) and it is a powerful thinking tool (as well as, occasionally, a cloak to disguise the absence of thought).

Although neurolinguists have found that language and thought are "subserved" (produced) by two distinct neural systems, there is no doubt that these systems are intimately related. (Section 3.1.) There is no speech understanding without ideation, and ideation is in turn greatly facilitated by language. The connections between the two neural systems are neural, hence material, not immaterial connections between two compartments of

an immaterial mind. Some such connections are newly acquired (as when we learn to name an object), others are lost (temporarily or forever), and still others change (as when we decide to give a new idea an old name, or an old idea a new name). The set of all such connections can be summarized into a code that pairs thoughts (and images, feelings, etc.) with sound sequences or visual images.

Thought can be encoded linguistically to the point that much of our speech and writing is made up of ready made phrases expressing thoughts we have had previously. This well known fact has suggested several hypotheses concerning the role of language in thinking and behaving. One is that language guides thought and behavior, so that the latter should be explained in terms of the former (Luria, 1961). However, developmental psychologists and psycholinguists (Piaget, 1955; Slobin, 1973; Clark, 1977) have found that the acquisition of language follows that of knowledge: in the child perception, imagination and ideation precede verbal expression. (For experimental evidence against Luria's hypothesis that language regulates behavior see Bloor (1977).) What is true is that (*a*) once a language has been acquired it can greatly facilitate the learning of new ideas expressed or expressible in the language; (*b*) whereas sometimes thought appears to come in the wake of language, usually it is facilitated by the latter, which on occasion replaces it; (*c*) no language can be learned except with the help of perception, feeling, imagination, and ideation.

A second conjecture is that the child is born with "the language of thought", or mentalese, which he then learns to translate into English, Sanskrit, Maya, or what have you. The way he learns to perform such translations is by making hypotheses, testing, and revising them. These hypotheses would be of the form "$L(x)$ is true (in language L) if, and only if, $M(x)$" where M is a predicate in the mentalese repertoire of the child (Fodor, 1975). Clearly, this extreme rationalist view renders the learning of new concepts impossible; also, it is at variance with neuroscience, which shows the cortex of the infant to be poorly organized. (See Smith Churchland (1978) for further criticisms.)

A third view concerning the importance of language is the Whorf–Sapir thesis. In its extreme form it states that language is a world view imposing upon the speaker a particular way of looking at the world and therefore of perceiving and understanding it (Whorf, 1956). This is a form of idealism close to Kant's: it may be called *linguistic idealism* (Ferrater Mora, 1967), and it happens to be false. The language of every society is part of its culture and is determined by its overall level of advancement, in particular its

fund of knowledge. A language is not a point of view, let alone a world view: it is subject-matter and ideologically neutral—so much so that one and the same language can serve to express mutually incompatible world views.

If a language lacks a word for the concept of atom, or of theory, this fact may indicate that it belongs to a culture whose members do not know about atoms or theories. But surely some of these people can learn to think and talk about such initially alien objects if they care to. As a matter of fact millions of Indians, Japanese and Chinese have assimilated much of Western culture and many of them have made important contributions to it—to the point that it is no longer correct to speak of 'Western' culture. What used to be given this name has become universal advanced culture, and its achievements can be communicated in many natural languages enriched with mathematics. Whatever expressive shortcomings a natural language may have are culturally conditioned and can therefore be overcome for, by and large, life molds language rather than the other way round. In particular, social status and geographical region determine the dialect the speaker will learn, not conversely. (See Trudgill, 1974.) In sum, the Whorf–Sapir thesis is false. (For further criticisms see Rosch (1977).)

A fourth well known thesis concerning the relation between language and reality is that language pictures the world: every sentence in a language would mirror a fact (Wittgenstein, 1922). A first objection to this thesis is that theories enriched with data, not languages, are supposed to represent the world. (Thus Maxwell's theory, not English, represents electromagnetic fields.) A second objection to the picture view of language is that it ignores thought. Actually we have to do with three different sets of object: facts in the external world, thoughts (brain processes), and linguistic processes (e.g. utterances): recall Vol. 1, Ch. 1, Section 3.2. And the correspondences between these sets are far from simple. In some cases a given external fact elicits no thoughts at all, at other times a single thought, and sometimes a numerous family of thoughts. Whereas some of these are verbalized, others are not; and, when they are, there is rarely a unique way in which they can be expressed. Finally there are thoughts without correlates in the external world, as well as bits of language that correspond to no thoughts at all: see Figure 3.5.

All four above theses concerning the relation between language and knowledge are false and they can be traced back to the theological assumption that wisdom is locked in Scriptures, or writings of alleged divine inspiration. This being so, wisdom can be ferreted out of words

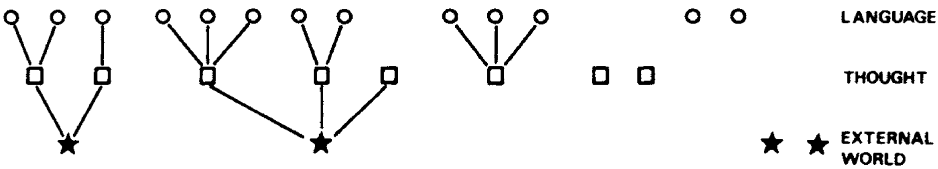


Fig. 3.5. Language, thought, and the external world. Some external world facts are not thought about, hence not described in any language, and some linguistic signs do not stand for any thoughts.

alone, be it by interpreting them correctly (hermeneutics) or by analyzing them (linguistic philosophy). Modern science rejected this approach and turned instead to experiment, theory and action. The later Wittgenstein and his disciples, as well as the German hermeneutic movement, have resurrected the biblical myth that in the beginning was the word—a counter-revolution in philosophy.

Language is both a brain tool and a social tool: it is a means for acquiring or creating knowledge as well as for transmitting it. To be sure not all knowledge is suitably communicated by word: thus motor skills and consequently much artisanal know-how is best learned by imitation. On the other hand all conceptual knowledge is communicable and thus socializable. Hence every major change in the mode of communication is bound to exert a deep influence upon the creation and diffusion of knowledge. In particular writing altered radically the mode of storing and sharing knowledge, as well as the mode of thinking: it made it possible to examine discourse in a far easier and detached way, namely by relying on visual perception. "No longer did the problem of memory storage dominate man's intellectual life; the human mind was freed to study static 'text' (rather than be limited by participation in the dynamic 'utterance'), a process that enabled man to stand back from his creation and examine it in a more abstract, generalized, and 'rational' way" (Goody, 1977, p. 37).

According to Comte every society is bound to go through three stages of development: religion, metaphysics, and science. Others have spoken of primitive and archaic knowledge, philosophy, and science. Finally others have distinguished the following stages: preliterate, literate but pre-scientific, and scientific. We may adopt the latter periodization, for literacy is a watershed: before it, knowledge could not accumulate beyond the storage capacity of the individual, and it could not be examined in a detached manner as if it were a material object. Besides, only a literate community can build public knowledge stores as well as collective cognitive

enterprises characterized not only by their bulk but also by their variety and pace. And only literate peoples can take advantage of the division of labor between those who produce, utilize or diffuse knowledge, on the one hand, and those who handle its externalizations—librarians, book printers and sellers, telecommunications workers, computer programmers, etc.

Literacy was a momentous achievement and yet, like all progress, it had some ambivalent and even some negative effects. Its ambivalent effects were these: (*a*) it gave rise to a new caste, that of scribes, which was (and is) usually in the employment and service of the powers that be—though occasionally some members of it turn against those powers; (*b*) it gave rulers an additional instrument of administration and oppression—but one which could also be turned against them. Literacy had also some negative effects: (*a*) it facilitated the defense of the establishment and its orthodoxy by introducing the categories of holy writ (or body of untouchable belief) and or legal code (or body of rules of conduct) that survived individuals and seemed to have lives of their own above the short-lived and miserable existence of mere humans; (*b*) it reinforced psychophysical dualism, i.e. the idea that thoughts can be detached from thinkers and subsist happily ever after in a world of their own (Plato's, Hegel's, or Popper's realm of ideas). In this respect literacy and, in general, language, is like any other instrument: it can be used for good or for evil.

4. Concluding remarks

Learning can be direct (personal) or, by means of communication, indirect (from others). Direct learning can be from experience and, in the higher animals, also from cogitation. Indirect learning can be by imitation or by instruction—or so-called 'self-instruction' with the help of books, journals, or computers. However, in social animals even direct learning, though a brain process, is done in society and largely through social intercourse.

Not only is all human learning social, but all social relations among humans are accompanied by communication of some sort or other. Therefore the study of cognition and language must not disregard their social matrix, and the study of social structure must not underrate communication. This does not entail that we are to blow up the sociality of learning or the informational aspect of sociality. We should not forget that only individuals (living in society) can do the learning, and that all means of communication are primarily means for transmitting knowledge or influencing the behavior of individuals. This caution is in order in view of

the contemporary tendency to reduce knowledge, and even social relations, to information flows. There is no information flow unless there is either knowledge to be obtained or transmitted, or action to be elicited or prevented.

Mankind has evolved the most expressive and precise means of communication known to us, namely language. This is not only our most sophisticated mode of social behavior and our paramount means of communication but also a powerful thinking tool, though no substitute for thought and not more valuable than images. Though indispensable to circulate knowledge and extremely useful as a thinking tool, language must not be mistaken for knowledge, just as money must not be confused with production and trade.

This closes our preliminary discussion of the acquisition, use and diffusion of knowledge. Henceforth we shall study particular modes of cognition, in the first place perception.

PART II

PERCEIVING AND THINKING

CHAPTER 4

PERCEIVING

Perception is the basic mode of cognition. Most of the animals capable of knowing get by with only perceptual knowledge. The higher vertebrates also attain conceptual knowledge, which enriches, corrects, and guides perception as well as action.

The best known view on perception is the so-called causal theory of perception, adopted by most empiricist philosophers (in particular Locke (1689)) and all behaviorists (in particular Pavlov (1955)). According to it all our perceptions are caused by external events and more or less mirror the latter. This view is a hypothesis rather than a theory (or hypothetico-deductive system), and it is only partially true. First, it ignores hallucination and illusion. Second, it ignores inhibition, which now obliterates, now enhances the effect of sensory stimulation. Third, it ignores that our expectations, values and hypotheses condition our perceptions. (We indulge not only in wishful thinking but also in wishful seeing.) We need then a truer account of perception, one making room for the creativity of the brain and, in particular, for the intertwining of perception and conception.

The causal theory of perception has been a constant concomitant of epistemological realism. This does not entail that, by correcting it, we have to give up realism. Even though our nervous systems are not passive recorders of environmental events, they do strive to perceive them correctly, and occasionally they succeed in doing so with the help of conception and action and, in recent times, also with the help of artificial sensors and recorders. We do not construct reality every time we look at or think of it. Le Roy was mistaken to assert that "*Le savant crée le fait*", and so are the anthropologists and sociologists of knowledge who unwittingly follow him believing that they are being faithful to Kuhn's (1962) somewhat ambiguous statements. The student of facts is given not only her brain and the cultural tradition into which she is born but also the very object of her study, namely the world—the only foundation of knowledge. What she is not given is a full understanding and control of her environment. But in a democratic society she is expected to learn something about her surroundings and herself, to make some contribution to the

understanding of the world, and to have a share in the control of it.

Perception occurs both at the beginning and at the end of every cognitive process concerning reality: it supplies facts to be accounted for as well as facts testing our accounts of the former. In this sense perception is the alpha and the omega of cognition. But by the same token it does not exhaust the entire alphabet: the remaining letters symbolize so many conceptual operations. One such conceptual operation is the explicit formulation of the problem posed by the initial perceptual data—e.g. What is that?, or Why is that so? Another is the formation of tentative hypotheses designed to solve such a problem. The test of such hypotheses joins conception with perception. Cognitive success lies then not in reducing conception to perception (radical empiricism) or in discarding the latter (radical rationalism) but in the alternation and confrontation of perception and conception.

1. From sensing to perceiving

1.1. *Sensing*

All organisms detect some environmental changes, but only the animals endowed with special detectors, namely sense organs, may be said to sense them. Thus whereas the octopus can see its immediate environment and discriminate among its components, presumably the sea urchin, which lacks sense organs, only detects changes in stimuli (von Uexküll, 1921).

Sensation is the specific function or activity of neurosensors such as the retina and the taste buds. Neurosensors are ordinarily multicellular systems specialized in detecting stimuli of certain kinds coming from the environment or from other parts of the animal. Like microscopes and photovoltaic cells, neurosensors filter out some (actually most) of the stimuli impinging upon them, and they accentuate the spatial heterogeneities and temporal changes in the environment, i.e. they enhance contrasts. For this reason they are similar to tuned filters (Ratliff, 1965, 1971).

Both the filtering out of stimuli and the enhancement of contrasts are results of inhibition. Thus when looking at a thing our success in seeing it depends not only upon the illumination of the external object but also upon the inhibition of the neurons in the visual cortex surrounding those that have been activated. (Thus, as Leonardo noted 500 years ago, when looking at a printed page you see all the printed symbols at a glance but, in

order to make out the words, you must focus on a small region of the page. In modern terms, in order to perceive what you detect you must inhibit your detectors.) The same holds for the other senses, in particular hearing and touch. This feature, lateral inhibition, is a peculiar property of the nervous tissue totally ignored by S-R psychology. It had been discovered by Mach one century ago, later forgotten, and rediscovered by Békesy (1967). From here on the minimal possible psychology is S-I-R, not S-R.

There is agreement that our sense organs are adaptations, i.e. results of genic variations followed by natural selection. However, such adaptation has not consisted in a gradual improvement of all the senses. Thus it is likely that our remote ancestors were capable of hearing, tasting and smelling better than we do, but we seem to be able to see better than they did. In addition to evolution we must take development and training into consideration. Thus a child's sensory discrimination is poorer than a normal adult's, and that of a technician or an experimental scientist (in particular an optician) is far superior to that of a layman.

Even so, our best senses have well known limitations. For example, if a thing moves very fast we see it as blurred or even not at all; and if a multicolored disk rotates rapidly we see only a homogeneous grey patch. And, no matter what the sensory attribute may be, we cannot distinguish without confusion more than about six different stimuli—e.g. visual position, or pitch, or salinity. In other words, our channel capacity is rather poor and is fairly uniform over the entire sensory domain (Miller, 1967). However, sensory discrimination can be educated to attain a remarkable sensitivity in musicians, microscopists, color experts, wine tasters, and other sensory specialists.

The phenomena of filtration, contrast enhancement, and confusion (poor discrimination) show that the vertebrate nervous system is not so much an information processor as a sink. When impinging on the central nervous system some stimulation gets lost in the following ways: (*a*) all sense organs are highly selective filters (e.g. we do not see infrared radiation or hear ultrasound); (*b*) the excitation reaching a point in the nervous system is not propagated but circumscribed (lateral inhibition); (*c*) excitations are usually dampened; (*d*) the various internal and external stimuli often cancel each other (compensation). All this shows that the nervous system is not at the mercy of its external or even internal environment. It is permanently active and it can do its job because of the above properties. In fact filtering cuts out much disturbing excitation, such as the infrared waves generated by our own body. Inhibition improves the

signal/noise ratio (e.g. yielding enhanced contrast). Damping protects us from incessant environmental shelling. And compensation discards much useless (together with some useful) information. By all these means the nervous system avoids information overload.

Accordingly the conception of the nervous system as an input-output box describable by a simple systems theory, such as information theory, is basically mistaken. Stimuli do not always excite, and those which do, do not always produce responses. On the contrary, all stimuli produce some inhibitions, and in the higher vertebrates much neural activity has no behavioral issue. (The typical mental activity is of this kind.) Therefore the S-R model is basically false. Worse, it has inhibited the investigation of inhibition and of spontaneous neural activity.

A similar criticism is applicable to the so-called ecological theory of perception (Gibson, 1966, 1979). According to this view, information resides in the environment: the organism picks it up ready made, processes it, and uses it, in analogy with the way it processes food. A first objection to this view is that information does not lie around us: we have to produce it, either by transforming sensation into perception, or by generating it internally (in ideation). Even the information-carrying signals impinging upon our neurosensors, such as the images on a TV screen, have to be "read" by the nervous system before they become information. Secondly, ambiguous figures, such as the Necker cube or some of Escher's *trompe l'oeils*, elicit mutually incompatible perceptions: they do not yield unambiguous information. The subject constructs alternatively mutually incompatible information packages out of one and the same stimulus. Thirdly, the subject, far from being a virgin film, has certain schemata (Bartlett, Piaget, Neisser) that allow her to perceive certain things rather than others, as well as to explore her environment searching for certain stimuli (e.g. moving objects) in preference to others.

In sum, the animal detects (some) environmental changes impinging upon it (i.e. it senses stimuli) and perceives ("interprets") some of them: it never picks up ready made information lying about it. Information about a stimulus equals perception of the stimulus. An unperceived stimulus carries no information; an incorrectly perceived stimulus carries wrong information; and an ambiguous stimulus elicits the generation of mutually incompatible pieces of information. In all such cases *information is generated by the organism* not just received and processed by it. Hence we should speak of *stimulus pickup and transduction* instead of *information pickup*, and we should drop the simplistic assumption that information

theory is the answer to the problem of perception. But we are encroaching on the subject of the next subsection.

To conclude this subsection. The neurosensors supply the raw material of our knowledge of the external and internal worlds. Without them we would know nothing, with them alone we would know nothing either. This is not philosophical dogma but a result of empirical research. In fact (*a*) subjects cut off from the world, in experiments on sensory deprivation, become spatially and temporally confused, and they start to hallucinate; and (*b*) subjects with their neurosensors intact but damages in the sensory areas of their brain cortex are incapable of processing the incoming stimuli: they can detect but not perceive (agnosia).

1.2. *Perceiving*

Physics and chemistry explain what happens in the eye from the moment light enters into it till it strikes the rods or cones in the retina. (Hearing is parallel.) From then on neurophysiology and psychobiology must take over. Thus the containment of the excitation, i.e. the inhibition of the cells surrounding those that are excited, is a typically biological process. The continuity of the visual world despite our blinking every few seconds may be an effect of a neural inhibition generated by the brain, that diminishes the effect of the blinks (Volkmann *et al.*, 1980). The discarding of small differences (as in pattern recognition) is the job of specialized (feature) cells in the primary visual cortex. And our perceiving distant objects almost as large as close ones, despite the fact that they project a smaller image on to our retina, is presumably a result of the activity of large neuronal systems in the secondary or tertiary visual cortex. All such phenomena illustrate the difference between sensation (or recording) and perception (or elaboration of sensation).

Whereas sensation is localized in the neurosensors, perception is distributed throughout large regions of the brain. Recently developed techniques for imaging and measuring local cerebral blood flow, metabolism, and other processes, allow neuropsychologists to locate the areas of the brain that are activated by a specific sensory stimulus. (Cf. Phelps *et al.*, 1981; Greenberg *et al.*, 1981.) One result is that the images of the brains of subjects given somatosensory, visual, or auditory stimulation are quite different. Another is that they have a significant overlap, suggesting that some neuronal systems are activated by sensory stimuli of any type. This may explain why our perceptions are strongly colored by our memories and

expectations and why they are accompanied by some movement or other.

We have no direct knowledge of the external world, for what is sensed is some event at the tip of a sensory nerve. The attribution of this even to an environmental item is a brain construction—either a perception or a hypothesis—and one that is sometimes correct and at other times incorrect. Thus a dog wounded by a stick does not snarl at its wound but at its attacker. We share with many animals "the power of imputing otherness and beyondness" (Lovejoy, 1929). The constructive nature of perception is further shown by the fact that we do not have to see a familiar thing in its entirety to recognize it: seeing only part of it, at a given angle, often suffices. (If it did not we would never recognize anything, for we cannot see any thing in its entirety.) In other words, neurosensors sample (and analyze) clusters of features, which are completed and integrated by the brain during perception.

Perception integrates sensation, is accompanied by behavior (e.g. eye movement during scanning), and is guided or misguided by conception. Indeed it has been shown experimentally that, if motor activity is impeded, e.g. by immobilizing the subject's eyes or limbs, then visual perception stops altogether. And if a subject wears a prism that reverses right and left, or top and down, and moves around, eventually the world "rights itself" for him: his plastic brain establishes new connections. In short, the way we perceive depends on the way we behave and conversely. (Cf. Taylor, 1962.)

Likewise what we perceive depends not only upon the stimulus object but also upon our knowledge and expectations. In short, in all likelihood there is no pure perception untainted by ideation and behavior. This result of recent physiological psychology had been anticipated by a number of philosophers. In particular Bacon, Kant and Whewell held that the way we perceive an external object is partly determined by our knowledge of it. Thus when I report that I am seeing Venus, I am drawing on my fund of knowledge: another person presented with the same celestial object might claim that he sees a star, a UFO, or an angel. This being so, it is very likely that our ancestors, even the recent ones, did not perceive the world exactly as we do. However, it was roughly the same world as ours.

Consider listening to speech or reading. Normally "concrete" words, such as 'run', trigger images, whereas "abstract" words, such as 'commutativity', trigger concepts or symbolic diagrams representing concepts. And when hearing or reading a new word either we form neither image nor concept, or we guess the correct image or concept from the function the word discharges in the sentence: i.e. we conjecture its meaning from certain

contextual clues. In short, speech perception and reading are highly constructive processes involving sensing, perceiving, imagining, remembering, conjecturing, and expecting. (See Carterette and Friedman (Eds.), 1976.) The reading of musical scores, diagrams, graphs, blueprints, and the like, must be parallel.

Percepts do not appear automatically and immediately upon sensory stimulation, but are constructed or synthesized (Selfridge and Neisser, 1960). This constructive view of perception has been confirmed by neurophysiology. In fact, we know from the work of Hubel and Wiesel (1962), Evans and Whitfield (1964), and others, that certain cortical cells specialize in recording single features (e.g. edge, inclination, or color), and that their output is synthesized by further neural systems. We also know that there are no isolated neural systems, so it is conceivable that the activity of the perceptual synthesizers is controlled by memory, expectation, and ideation. See Figure 4.1.

Throughout the history of philosophy and psychology two major views on perception have clashed: those on perception as recording (or passive models) and as constructing (or active models). Naive realism and empiricism have held the former, whereas idealism has held the latter. The passive models are the slate and the filter: according to them all the nervous

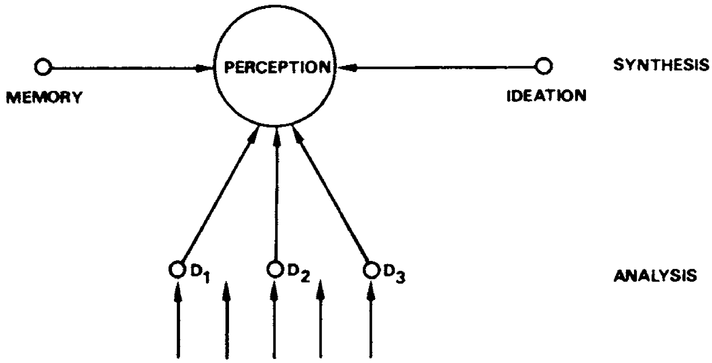


Fig. 4.1. Cells D_1, D_2 and D_3, of different kinds, respond selectively to different features of one and the same stimulus, and their outputs are synthesized by a neuronal system (the perceptor). Some of the features of the stimulus are either not detected at all or suppressed. The synthesis is not a neutral summation but is influenced by the current state of the whole brain, in particular its memories and expectations. Thus when we say "I know this girl" when meeting a particular girl, we mean that we have recognized her: our perception of her is "colored" by our memories of her and perhaps even our expectations about her.

system does is to record and select the incoming stimuli. A deterministic automaton with or without output would be a good mathematical representation of such models.

Obviously, no passive model harmonizes with neuroscience, in particular the findings about lateral inhibition and the spontaneous ongoing activity of neurons and neuronal systems, in particular the integration of the outputs of specialized detectors (recall Figure 4.1). These findings support on the other hand some active or constructive model, according to which perception is a brain activity resulting from the joint action of input signals and internal processes. As a matter of fact it has been known since Helmholtz (1879) that the senses, far from giving us a *picture* of the world, give us only *signs* of it, which signs must be "interpreted" by the brain to become cognitive items. In short, we do not just copy or reflect but construct. However, this does not endorse idealism: we do not construct the world but only map it.

Every animal is equipped not only with a number of sensory systems but also with certain predispositions or propensities to use such systems in certain ways rather than others. Thus we are predisposed to paying attention to moving objects and shrill sounds; to listen to and imitate human speech in certain languages rather than others (similarly most song birds imitate the song of their conspecifics even when they are exposed to the song of birds belonging to other species, or to music); we notice items of certain kinds more readily than others (e.g. expecting parents tend to see pregnant women and babies everywhere), and so on. These perceptual propensities are inherited or acquired properties of our brain and they develop with it: they allow us to have certain experiences and are in turn modifiable by the latter. Psychologists call such predispositions perceptual *schemata* (Bartlett, 1932; Piaget, 1952; Neisser, 1976).

A perceptual schema is neither a thing (e.g. a neuronal system) nor a process in a thing. It is a cluster of properties of a neuronal system and, more particularly, a bundle of propensities for it to act in a certain manner in the presence of external or internal stimuli of a certain kind. Presumably such propensities can be quantitated as probabilities, so that all models of perception should be stochastic models. However, this is just conjecture: we still do not have adequate models allowing us to explain how a person in a certain state sees a chair as a piece of furniture, and when in a different state as a weapon or as a piece of firewood. The most sophisticated models of perception, those of Hoffman (1966), are not stochastic but use Lie groups, and they do not involve any neuroscience. A characteristic

hypothesis of such models is that a global pattern is the result of an iteration of Lie operators—rather than, say, the activity of a neuronal system activated by a number of sensory analyzers or detectors.

Adequate models of perception can be produced only by neuropsychology (or physiological psychology): they will not come from pure psychology, let alone from philosophical psychology, which has been at it for over two millennia without ever getting hold of that which does the perceiving, namely the central nervous system. A good example of the failure of armchair speculation on perception is the "new theory of perception" advanced by Brown (1977). This view is claimed to elucidate Hanson's idea that scientific percepts are theory-laden. (Actually this idea goes back to Whewell, who took it from Kant, who knew about Bacon's *idola*.) The gist of this view is that the distinction between *seeing* (e.g. a shadow) and *seeing that* (e.g. that the shadow has been projected by a person) is that in the second case the observer takes in "the meaning of the observed situation" (Brown, 1977, p. 87). The author calls this 'significant perception' in contrast with sheer perception (actually sensation). But of course there is no *theory* proper (i.e. hypothetical-deductive system) here. There is not even analysis, for Brown does not elucidate the key notion of "meaning of a situation". And there is no novelty, for the whole thing boils down to the old finding that we perceive what we are prepared to perceive. The challenging task is not to use ordinary language expressions to describe once more the difference between untutored sensation and educated perception, but to build a psychobiological theory couched in mathematical language.

Understandably, most of the research in perception has been concerned with vision. (If we dwelt in deep caves we would presumably have specialized in tactile and auditory perception.) However, not all perception is of some external object: there is also perception elicited by bodily processes. Thus we perceive the pressure of one hand rubbing the other, and we perceive (painfully) the thorn in the toe or the ulcer in the duodenum. Besides, the brain can generate perceptions that have no stimulus object, whether external or internal: such are hallucinations and imaginations. (The only difference between these two is that the former falsely imputes some external stimulus object.) We can imagine pictures, sounds, bodily processes, and feelings. We can do so at any time, while awake or in sleep, working or daydreaming, in normal circumstances or under the action of drugs or of hypnosis. (There is no such thing as a hypnotic trance: the good hypnotic subject is highly suggestible and

imagines whatever is suggested by the hypnotizer: Barber and Wilson, 1979.)

Imagination is a mode of cognition without which nothing great is ever achieved in any field. True, the products of imagination must be checked for truth or efficiency unless we are bent on creating a work of art. But the same holds for ordinary perception and conception as well: we are not to trust any single cognitive operation. Moreover such checking is never automatic and infallible: it takes a lot of imagination to produce firm evidence for or against the works of imagination. Only the empiricist obsession with empirical data could make us forget for a moment that the most valuable data are those produced with the help of imaginative means such as sophisticated experiments inspired by sophisticated theories. Scientists and technologists do not shun imagination: they cultivate it and keep it under control.

Imagination takes part also in the intuitive or ordinary knowledge of other minds known as empathy. Empathic understanding, which we share with many animals, is nothing but imagining what goes on in someone else's brain, or "putting oneself in the other's shoes". It is compassion, coperception and coideation. Empathy uses (ambiguous) clues and (fallible) conjectures relating overt behavior to mental states, e.g. the shedding of tears to sadness, and the shuffling of feet to embarrassment. Like every mode of perception, empathy is fallible: the illusion of infallibility comes from our noting its hits and disregarding its misses. "Lovers, who are notoriously supposed to be empathic, as frequently seem to misunderstand as understand each other's motives and acts; and what they normally do, perhaps, is wilfully misunderstand their existing differences, so that they create an illusion of unanimity of opinion and feeling" (Ellis, 1963, p. 119).

What about space and time: are they perceptible? Although many psychologists write about the perceptions of space and time, these are imperceptible: only (some) events are perceptible (under special circumstances). What we can perceive is not space but events that hold spatial relations; nor do we perceive time but events in temporal relations. Thus we hear one noise to the right of another, and perceive the former before the latter, but "to the right" is just as imperceptible as "before". In general, we perceive neither things nor properties nor relations: we perceive only some events, which are changes in properties of things or relations among things. If there were changeless things we would not notice them.

Finally, can we perceive external objects without the help of the senses:

i.e. is there extrasensory perception (ESP)? In the context of traditional psychology, which ignores the nervous system, this question must be regarded as open, for one might hope (or fear) to get eventually some firm evidence (other than statistical mistake or deception) for the hypothesis of ESP. Likewise before the nature of meteorological events was understood one could hope or fear for a storm to strike out of the blue. We now know that the perception of a stimulus object (unlike imagination and hallucination) is the terminus of a process that starts at some nerve ending or other, and ends up in a secondary or tertiary sensory area in the cortex. This being so, clairvoyance and other alleged ESP phenomena are as illusory as running without limbs or digesting without guts.

It is a firm finding of physiological psychology, and a principle of our epistemology, that all external perception is mediated by some sensory system. Put negatively: there is no ESP and there can be none. In particular there is no direct communication among animals at a distance: all communication proceeds *via* physical signals (light, sound, touch, etc.). And what mind reading there is consists in perceiving and interpreting the movements or signals produced by another animal. Thus you are reading my mind as you read these lines.

2. Observing

2.1. *Direct Observation*

There is an important difference between spontaneous and directed perception, or sheer perception and observation. In the former case we look *at* things, in the latter we look *for* them. Directed perception, or observation, is purposive, active, and deliberately selective. A dog and a human being watching the passing scene are onlookers; a dog looking for a bone, and a human being looking for a friend, are observers. The difference is one of psychological "set" (*Einstellung*): different neuronal systems are presumably involved.

Because observation is purposive and selective perception, it can be educated. (Recall Leonardo's injunction to his students: *Saper vedere.*) In fact the training of a young hawk, a tiger cub, or a human infant, are largely education in observation skills: in focusing on certain things or traits while disregarding the rest, or enhancing the effect of certain stimuli while inhibiting others. (Békesy (1968) stated that "a good observer is an observer who learns to inhibit". But so is an observer incapable of

perceiving due to bias or to unwillingness to face reality. Thus Tolstoy's Karenin "was unwilling to see anything and so saw nothing".) There is little doubt that our children, who are exposed from birth to a ceaseless stream of pictures and noises, are better lookers but worse listeners than our fathers. In the case of humans this education is not only sensory-motor but also intellectual. We perceive almost exclusively that which we have been trained to perceive (including self-training). If our brain has been soaked since childhood with witches, devils and conspirators, we are bound to look for them and likely to find them. If on the other hand we have had the benefit of a scientific or technological education we shall tend to look only for concrete things that can be accounted for in naturalistic terms.

Two persons, a farmer and a scientist, see a brilliant thing now hovering above them, now moving swiftly away, and neither can identify it immediately. If the farmer has heard about angels or UFOs (first "seen" in 1947), he may believe he is seeing one of either. The scientist, on the other hand, is likely to identify the object as an aircraft, a plasma, or Venus. Philosophical subjectivists notwithstanding, both subjects deal with the same object perceived differently. This holds not only for brilliant flying objects but for almost everything: people with different backgrounds, cognitive attitudes, and value systems, are bound to see the (only) world there is in more or less different ways.

Some philosophers, in particular Whewell (1847), Duhem (1914), and Hanson (1958), have stated that experiences like the above show that observation, far from being neutral or uncontaminated by theory, is penetrated by it. (Hanson introduced the expression *theory-laden*, which is incorrect, for most observers do not know any theories proper. The expression *hypothesis-driven* seems preferable.) There is a grain of truth in this: we do tend to observe what we expect to perceive, we often do miss what we do not anticipate to perceive, and all our expectations are determined by our fund of knowledge (and superstition) as well as by our value system (or interests). However, that grain of truth should not be puffed up to the point of epistemological relativism and subjectivism. The observations on the brilliant flying object reported by the farmer and the scientist are not equally correct: they are mutually incompatible, both can be checked for objective truth, and it so happens that the checks favor the scientific version.

True, observation is hypothesis-driven, particularly in the sciences, because it is selective. The myth of immaculate observation is all but dead, even in predominantly observational disciplines such as ethology (Hinde,

1970; Beer, 1973). But, if scientific, rival hypotheses are checkable. And, once checked (against data or theories or both), they are bound to be shown unequal in worth (truth value, generality, or depth). The same holds for data: many of them can be replicated, and all of them are (in principle) checkable by different means or methods. Consequently even if all of them are hypothesis-driven, many of them are *invariant with respect to changes in the driving hypothesis*. So, such data are in effect uncontaminated by any particular hypotheses. This is the case with many data stemming from direct observation, i.e. observation unaided by any instruments, such as those gathered by a census bureau. On the other hand indirect observation, i.e. observation conducted with the help of scientific instruments, is not only driven by hypotheses but is also hypothesis-dependent, i.e. its worth depends in part upon the hypotheses involved in the design and operation of such instruments. We shall return to this in the next subsection.

What is the epistemic value of observation? According to empiricism all and only observation supplies genuine knowledge: i.e. observation is necessary and sufficient for it. Epistemological analysis shows that observation, though necessary, is not sufficient, for all observation is hypothesis-driven, and all indirect observation is hypothesis-dependent. Thus finding a fossil bone is not, in itself, a scientific discovery: many people had found, and discarded or misinterpreted, fossil bones before evolutionary biology made it possible to "interpret" them as remains of organisms that lived long ago. An empirical scientific discovery is a finding together with a theoretically plausible account of it. Hence two rival accounts of one and the same finding must rate as different discoveries—at most one of which can be correct. Thus Columbus and Vespucci, though interpreting roughly the same geographical findings, made different discoveries: in fact only Amerigo Vespucci discovered America. Likewise although the ancient Greeks knew about amber and loadstone, they did not discover electricity or magnetism; and the discoverers of DNA did not realize that it is the genetic material, and so did not discover that DNA is the carrier of heredity.

Observation, though insufficient, is necessary for genuine knowledge of fact. True, we cannot observe directly neutrinos or synaptic boutons, any more than live dinosaurs or prehistoric events. But we do make certain observations of related things that motivate and check the hypothesis of the current or past existence of unobservables. On the other hand there is no observational evidence for the hypothesis that there are disembodied entities, such as ghosts and ideas in themselves: hence this hypothesis does

not count as genuine knowledge. The need for observation in gaining factual knowledge is a tacit postulate of science and technology. (Even the highest level scientific theories, such as quantum mechanics, contain constants or parameters whose values can be known only through measurement. And all measurement involves, at some point or other, direct observation.) Therefore we incorporate that principle into our epistemology: *All knowledge of factual matters consists of, or involves, some direct observation.*

The historical sciences, such as cosmogony and geology, archaeology and human history, are not exempt from this condition. True, historians cannot observe events in the remote past: most of their observations are limited to remains, such as tools and written documents. They must conjecture the possible use of tools and, if they proceed scientifically, they will make replicas and try them out; and they must be able to interpret the documents and try and check them with other findings. In all cases, regardless of the amount of speculation historians may embark on, they have to make some contact with observation if they wish to substantiate their hypotheses. The same holds for all the other historical sciences, from cosmogony to evolutionary biology: there is always some observation involved in them, whether at the beginning or in the middle or at the end of a research cycle. A field of inquiry that involves no observation at all is either strictly formal (i.e. logical, mathematical or semantical), or purely speculative—and therefore nonscientific.

2.2. *Indirect Observation*

Not all facts are directly observable. By far the greatest number are not: think of electrons or quasars, of genes or nations, of atomic collisions or historical events, of extinct organisms or neural processes. All such facts go beyond appearances and therefore constitute a glaring refutation of phenomenalist epistemologies and metaphysics.

Suspecting that reality encompasses more than appearance, scientists have devised literally thousands of means for extending the reach of our senses. Strictly speaking only a few such means—e.g. the scale, the microscope and the telescope—extend the scope of the human senses. Others, such as the magnetometer and the pH meter, the voltmeter and the interferometer, the Geiger counter and the EEG, are not extrasensitive eyes, ears, tongues, or skins, but tools without biological precursors. However, all of them supply directly observable facts, such as clicks,

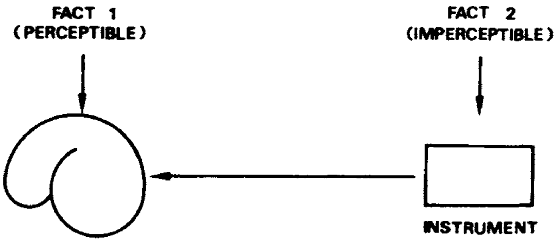


Fig. 4.2. Direct and indirect observation.

coincidences between pointers and divisions on a dial, or computer printouts. The mode of inquiry employing such means may be called *indirect observation*. See Figure 4.2.

A short and haphazard list of such means should show how much farther than our senses they go, as well as how hypothesis-dependent they are. Radiotelescopes are indispensable to study radiostars. Astronomers hope (or dread) to discover black holes either by looking for very luminous stars (binary pulsars?) that seem to revolve around nothing, or by observing the radiation emitted by objects in the process of being swallowed by black holes. Neutrino detectors allow one to "look" into the interior of a star, where they are produced, for starlight comes mostly from the thin stellar atmosphere, and neutrinos can traverse it almost undisturbed. (Incidentally, such observations have recently upset the accepted theories of the structure of stars.) Field-ion microscopes and scanning transmission electron microscopy produce direct images of individual medium to heavy atoms and molecules. (The classical techniques supply only images of average structures.) And there are several techniques for detecting what goes on in the living human brain (e.g. microelectrodes and EEG) and even for imaging such processes—e.g. computer assisted tomography (CAT scan) and positron emission tomography (PET scan).

As with direct perception, instrument assisted detection is of events or changes, and the detection itself is a change in a thing, namely the detector. To be sure the thing detected may appear to be in a stationary state, as is the case with the printed page you are looking at right now. But this is only appearance: in actual fact this page, while legible, is being bombarded by myriads of photons, some of which are absorbed, the rest reflected, and all of which affect this sheet in some manner, e.g. by warming it up and speeding up some chemical reactions in it. In short, only events are detectable.

Every means of indirect observation is conceptual as well as material, for it involves not only a material instrument, such as a scale or an amplifier, but also a set of hypotheses employed in its very design and operation. Thus, the detection of radio waves employs electromagnetic theory; X-ray and neutron crystallography involve not only pieces of apparatus but also the theory of diffraction; electrophysiology uses not only electrodes and meters of various kinds but also scraps of physical and biological theories. In many cases the theory itself suggests the very existence of the thing to be detected as well as manners of detecting it. (Thus a certain technique for detecting the elusive neutrino is based on calculations showing that, when the nucleus of gallium 71 absorbs a neutrino, it turns into the nucleus of germanium 71, a different element.)

Without such theoretical components the raw data supplied by indirect observation would be unintelligible and therefore useless, for they bear no obvious relation to the facts that are being observed. Thus a layman would not know what to make of a sequence of clicks in a Geiger counter, or of an X-ray diffraction pattern, or of a collection of spots on a sheet of chromatographic paper. Every indirect observation requires an "interpretation" of the data, which "interpretation" consists of hypotheses about the nature of the events of interest and their interaction with the detector. This intervention of hypotheses in observation is not limited to natural science, medicine and engineering, but occurs in all fields of research. Thus sociologists use social indicators, and economists economic indicators, to detect changes in society, and prehistorians examine artifacts, bones, and inscriptions to "read" the evolution of mankind. In all such cases certain hypotheses—that will be called *indicator hypotheses*—link the available data with the events of interest. (More on indicators in Ch. 11, Section 2.2.) In short, every means of indirect observation, be it of microevents such as atomic collisions, or of macroevents such as economic crises, involves a theoretical component—or, as an empiricist would say, is "contaminated" by theory.

Indirect observation has advantages and disadvantages with respect to direct observation. The advantages are obvious: (*a*) it enables us to go far beyond appearance; (*b*) because it employs impersonal instruments free from conceptual or moral bias, and yielding data that can be perceived by all competent observers, it is less subjective than direct observation. The disadvantages are equally evident: (*a*) some instrumental probes disturb the probed object (particularly in microphysics), and all instruments are likely to produce "artifacts", i.e. effects of their own (e.g. halos and stains) that the observer may mistakenly impute to the observed object; (*b*) the

data resulting from indirect observation are just as fallible as their theoretical components. Not surprisingly, if one wishes to get more one must be prepared to pay more. However, the cost (i.e. error) may be lowered with improved instrumentation and theory.

As for bias caused by wishful thinking or dishonest intention, indirect observation is neither more nor less susceptible than direct observation. One can lie not only in reporting sightings with the naked eye but also with a camera—e.g. by changing the sequences of frames, or taking pictures from appropriate angles or under suitable illumination—or by splicing tapes or simply by selecting the favorable data and eliminating all the others. However, it is obvious that lying with cameras (or meters or computers) goes down better than without them. Hence the methodological and ethical precautions are more necessary in the laboratory than in the field or the street, the more so since all results of indirect observation must be doctored in some way or other.

Advanced science has little use for raw data: the latter must always be corrected for a number of effects, and sometimes (as in the case of outlying figures) thrown away altogether. The line between standard data correction and finagling, between honesty and dishonesty in reporting results of observation, is not always a clear one. Newton has been accused of smudging astronomical data and Mendel of ignoring certain results of his hybridization experiments. Morton has been shown to shift criteria, and omit or miscalculate some data on cranial capacity for harboring "an a priori conviction of racial ranking" (Gould, 1978); and Millikan discarded some values of the electric charge of the electron for failing to square with his hypothesis (since amply confirmed) that all electrons have the same charge (Holton, 1978). In all these cases preconceived ideas—sometimes true, at other times false—had distorted the presentation of data. Nor are such cases exceptional: they occur all the time in ordinary life and in the laboratory. Thus if we are used to seeing A's but not B's under certain circumstances, or if we assign A's a greater probability or utility than B's, we may miss the latter when they occur or confuse them with A's—or may be inclined to report only A's for fear of being taken for fools. (For an experimental study of biasing factors in discrimination problems, see Swets (1973).)

Not all observers are equivalent in all respects—*pace* Ziman (1979). Therefore we should not expect that different observers will always produce the same reports on one and the same group of facts. Two observers may be inequivalent because they are not in the same state of motion, or because they use different equipment, or operate the same equipment in different

ways, or under different circumstances. Or they may be inequivalent because they handle their data (e.g. round off, average, or discard figures) in different ways. Or because one of them is gifted or lucky enough to notice something missed by the other, or because one of them is wedded to a certain hypothesis whereas the other is indifferent to it. (Recent examples of the effect of theoretical bias have been seen in the search for gravitational waves, black holes, and quarks.)

Even the most competent and lucky of observers is faced with the inherent limitations of every means of observation. All optical instruments have a limited resolving power due to the diffraction of the impinging light. Electronic and protonic microscopes possess a far greater resolving power but have their own limitations. Astronomers are up against the limits of their instruments as well as of the observation sites: clouds and turbulent currents, not to speak of smog, will produce a diffuse and constantly fluctuating image because of the variation in the atmospheric refraction index. Obviously, such limits to the accuracy of observation are limits to knowledge. However, some such limits can be overcome in principle and possibly in practice. For example, telescopes installed in artificial satellites or on the Moon will be free from atmospheric nuisances, and all kinds of electric meters are essentially free from thermal noise when working at extremely low temperatures.

We summarize the above discussion on the reach and limits of observation in the following *postulate. For every fact f* (thing or event) physically accessible to cognitive subjects (*a*) *there is* (actually or potentially) *a means m of observation such that f has observable traits relative to m* (i.e. such that *m* helps detect some properties of *f*), *but (b) there is no means allowing one to observe all of the traits of f with complete accuracy.* In short, the means of observation can all be improved but never be rendered perfect. (*Notes*: (*a*) the physical accessibility condition excludes most events in the remote past, as well as all future events and those outside the observer's light cone; (*b*) the limits on observation have nothing to do with Heisenberg's indeterminacy inequalities, which only limit the applicability of the classical concept of a corpuscle.)

3. Mapping Reality

3.1. *Perceptual Mapping*

Perceptions, whether spontaneous or sought, do not pile up haphazardly but tend to get organized into systems. Such systems of percepts are called

cognitive maps or *internal representations*. Some of them represent our environment (*environmental maps*), others our own internal states (*body maps*). These maps are continually updated, they summarize our perceptual knowledge of our surroundings and of ourselves, and in turn they help us take care of ourselves, move around, and act: see Figure 4.3.

We do not know exactly what animal species are capable of forming perceptual maps. All we seem to know is that, in order to map or model its surroundings, an animal must possess a nervous system capable of (*a*) *analyzing* or breaking down into distinct units whatever it perceives, (*b*) *recognizing types* or kinds into which such units can be grouped, and (*c*) *organizing* what it perceives into a whole, i.e. synthesize or integrate the products of analysis and type recognition. Presumably, it takes a rather complex nervous system to accomplish all this—say, that of a fly at least—but we do not know exactly what it takes.

According to psychoneural dualism, cognitive maps are immaterial: they are "in the mind". So, by hypothesis subhuman animals would not be entitled to forming cognitive maps. On our biological perspective a cognitive map is a pattern of neural activity in perceptual and motor centers. In particular, it has been conjectured that the internal representation of an object seen by a normal human subject is a sort of ring formed by sensory memory traces and motor memory traces (Noton and Stark, 1971). As a subject sees a thing for the first time, she forms a scan

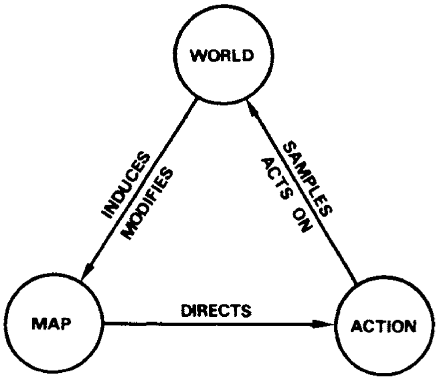


Fig. 4.3. Cyclic model of perception (adapted from Neisser, 1976, p. 112). External objects stimulate the formation of cognitive maps in the brain. These in turn guide locomotion, exploration, and other activities that modify the environment, which changes call for revisions in the cognitive maps, and so on. Caution: both maps and actions are part of the subject, and while both must be conceived as biological processes, the world is a thing.

path of that object. This scan path consists in a chain composed by a visual feature followed by the eye movement required to reach the next visual feature, and so on. When the subject encounters the same object again, she recognizes it by matching the incoming visual stimuli with the "feature ring". Thus once constructed each such ring guides the future perception of similar objects. Each subject constructs her own feature ring, which is fairly permanent. This hypothesis may not be accurate but it matches the available experimental data—particularly all the data concerning the strong linkage between perception and motor behavior—as well as the general psychobiological orientation. It is incompatible with psychoneural dualism—a definite virtue in our view—as well as with the Gestalt hypothesis that internal representations are built in one piece and instantly.

(One way of elucidating the general concept of a cognitive map is as a function from some events in the perceptual field into neural events. More precisely, consider a concrete thing m and take the set M of events happening in m at a certain level L of discrimination—e.g. the collection of all the photons of a certain frequency bouncing off m. Further, call M_L the subset of M formed by the events above a certain level L of discrimination. Then a *cognitive map* of the thing m is a function f from a subset P_L of M_L to a set N of neural events in some subject, i.e. $f: P_L \to N$, that preserves some of the relations and operations among members of P_L. A perfect cognitive map of m is of course one that (a) mirrors every event happening in m, i.e. for which $P_L = M_L$, and (b) preserves all of the relations and operations in M_L. We may safely assume that no such perfect map is possible.)

Obviously, the cognitive maps of different individuals, even of the same species, are not identical: each of us perceives herself and the world in her own way. Still, there must be commonalities due to our common ancestry and our shared environment. In general, the individuals of each animal species would seem to map their surroundings in a species-specific way: the octopus in one manner, the owl in another, and so on, as proposed by Uexküll (1921). (Actually Uexküll held that each animal *constructs* its own environment in a species-specific way. This is either sloppy language or subjective idealism. All we do is map our environment, and use such maps to navigate in it and modify it. More on this subject in Ch. 15, Section 2.2.)

In the recent psychological literature the expression 'cognitive map' is usually understood in the restricted sense of a *spatial* map or schema allowing an animal to orient itself in ordinary physical space. Thus O'Keefe and Nadel (1978) have conjectured that all mammals are born with a framework, located in the hippocampus, where they can store all the

information concerning places and spatial relations that they gather in the course of their lives. They further conjecture that, "When there is a mismatch between some sensory input to a place representation and the predicted sensory input, a set of *misplace* detectors is triggered, the output of which activates and directs the motor systems involved in exploration. The behavior is directed towards the incongruence and new information can be incorporated into the map as a result of it" (p. 94).

That some animals form spatial maps of their environment, and that such maps guide much of their behavior (approach, avoidance, exploration, etc.), seems beyond reasonable doubt. What may be disputed is the thesis of Descartes and Kant, revived by O'Keefe and Nadel (1978), that all mammals are born with a *Euclidean* framework. In fact it has been known for some decades that (*a*) none of the various perceptual (visual, auditory, tactile) geometry is Euclidean—e.g. some parallel straight lines are seen as curves, whereas some curves are perceived as straight parallel lines; (*b*) children learn rather slowly, when at all, the "facts" of Euclidean geometry. Luneburg (1947) had proposed that visual space has a constant negative curvature, i.e. is Lobachevskyan. Battro *et. al.* (1976) have challenged this thesis on experimental grounds, holding instead that visual space has a variable curvature: it is sometimes positive (elliptical geometry), sometimes negative (hyperbolic geometry), and occasionally null (Euclidean geometry). As for the development of geometrical ideas, we know now that newborns move their heads to locate sounds and sights, so they do have some inborn spatial framework. But we also know, from Piaget's studies,

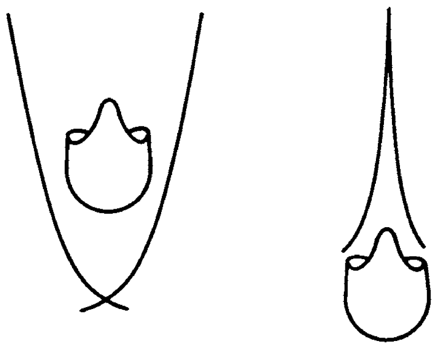


Fig. 4.4. Natural perspective is curvilinear. If asked to trace an avenue or parallel straight sides, the subject imagines curved lines. And if he looks at straight rail tracks he does not see them converging in straight lines but as branches of a hyperbola (Battro, 1979). Curvatures in drawing greatly exaggerated.

that this framework must be rather poor, for the triangle inequality does not appear normally before age 11, and even then only among half of the children (Piaget and Inhelder, 1967). Besides, our natural (untrained) perspective is not rectilinear but curvilinear (Battro, 1979): see Figure 4.4. In conclusion, we are not born Euclidean—or Riemannian or Lobachevskyan. If we did, geometry would not have been such a recent acquisition, and its teachers would have a far easier task.

3.2. *Appearance and Reality*

At birth we have only our perceptions to go by, so presumably we are born phenomenalists: only reflection upon our experience, together with education, turns us gradually into realists. At age 2 we may still reach for the Moon, but by age 4 we have already sobered up. The very young child presented with two things of the same volume but very different shapes, such as a ball and a pencil, tends to judge incorrectly their sizes: he sticks to appearances. Only the child who has attained the "concrete operational stage" tends to recognize the conservation of matter underneath changes in shape, color, and other appearances (Piaget, 1976a). The same holds for adult trained chimpanzees (Woodruff *et al.*, 1978).

We never see the world exactly as it is. For one thing we fail to perceive stimuli below a certain threshold and above a certain ceiling. For another we tend to inhibit most sensations. Thirdly, our bodily state changes from one moment to the next, so that a constant external stimulus is transduced now in one way, now in another. Fourthly, our neurosensors are such that "Humans do not experience the world as it is but as a power function of what it is" (Thompson, 1975, p. 239). This cryptic albeit true and profound statement may be understood as a verbalization of either Stevens' psychophysical law or Thouless' index. The former states that the perceived or apparent intensity of a sensory stimulus is a power function of the physical or objective stimulus. And the Thouless index relates the real and the perceived (apparent) sizes of a visual stimulus object, as well as the projection of the latter on the retina: see Figure 4.5. In sum, pure (sensory-motor) experience yields appearance not reality. If we wish to attain reality we must go beyond perception, into conception and action.

The perception of a fact is called a *phenomenon* or *appearance*. (In ordinary language 'phenomenon' is equated with 'fact': beware of the imprecisions of ordinary language.) There are imperceptible facts but there are no phenomena without sentient organisms. Appearance, then, is an

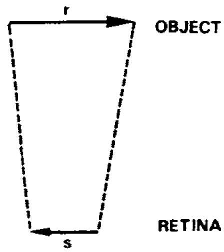


Fig. 4.5. The Thouless index: $i = \log(p/s)/\log(r/s)$, where p is the perceived or apparent size of the stimulus object of size r, whose projection on the retina is s. By algebraic manipulation we obtain $p = s(r/s)^i$. If $p = 4$, then $i = 1$; and if $p = s$, then $i = 0$.

evolutionary gain: it emerged together with the first animals equipped with nervous systems. Before them facts appeared to nobody: there was no appearance, there was only reality.

Phenomena are facts of a special kind, namely facts occurring in nervous systems. So, *phenomena are real.* Consequently there is no opposition between appearance and reality. My seeing the Moon larger on the horizon than overhead is a fact no less than the two positions of the Moon: only, the former is a perceptual, hence *subjective*, fact, whereas the latter are *objective* physical facts. There is then nothing wrong with admitting phenomena alongside nonphenomenal (or transphenomenal) facts. The opposition is not between appearance and reality but between subjective facts or accounts and objective facts or accounts.

From the fact that perception gives us only appearances, and the (false) hypothesis that perception is the only source of knowledge, some philosophers have concluded that there *are* only appearances, and others that only appearances can be *known*. The former (e.g. Berkeley) are *ontological phenomenalists*, the latter (e.g. Kant) are *epistemological phenomenalists*. Ontological phenomenalists hold that there is no world independent of the knowing subject: that to be is to perceive or to be perceived. Epistemological phenomenalists hold that, although there may be an autonomous world, i.e. a reality independent of the knowing subject, it cannot be known: at most we may construct fictitious worlds. Clearly, ontological phenomenalism entails epistemological phenomenalism.

A rigorous phenomenalist cannot go beyond appearances and must limit himself to describing or forecasting them: any attempt to understand them would lead him far beyond his own experiential horizon, into the wide world. Thus he will utter subject-centered sentences such as *Phenomenon x*

appears to subject y under circumstance z. To the realist this is just a truncation of the full sentence *Phenomenon x, appearing to subject y under circumstance z, is caused by (or is indicative of) noumenon w*. And, in addition to such mixed sentences involving objective as well as subjective items, the realist will construct strictly objective sentences such as *Fact x occurs under circumstance z*. Indeed, he will attempt to translate all subjective and mixed sentences into strictly objective ones. This does not entail ignoring subjectivity but putting it in a wider context as well as trying to explain it.

Needless to say, scientists and technologists, regardless of their explicit profession of philosophical faith, behave like realists not phenomenalists. Thus if somebody complains that he feels X, his physician will probably conjecture that his patient has (objectively) ailment Y, which manifests itself as symptom (phenomenon) X. And the biologist will try to explain this cause-symptom relation in physiological terms. Nor is this realistic attitude confined to science and technology: it also permeates the arts. Thus Leonardo and Michelangelo studied anatomy in order to understand what they saw and thus be able to reproduce it more faithfully. Those who do not know "which nerve or muscle is the cause of each movement" are bound to "draw their nude figures looking like wood, devoid of grace" (Leonardo, 1492 or 1515, No. 488).

Phenomenalism, in either version, has been an obstacle to the inquiry into the real world, for it discourages the hypothesizing of objective facts behind phenomena, and it encourages instead the formation of superficial (phenomenological, behavioristic, black-box) hypotheses, and even the survival of an anthropomorphic world view. (Phenomenalism is the last ditch stand of anthropomorphism.) Nevertheless, phenomenalism has been embraced by a number of influential thinkers after Berkeley and Kant, such as Mill, Avenarius, Mach, Whitehead and at one time Russell, Carnap, Ayer, and Goodman. Besides, it has made a decisive contribution to the methodology of behaviorism (Watson, 1925), to the operationalist philosophy of physics (Bridgman, 1927), and to the so-called Copenhagen interpretation of quantum mechanics.

Phenomenalism, in either its ontological or its epistemological version, is false. First, as noted long ago by Plato (*Theaetetus* 158) against Protagoras, our ability to distinguish dreams and illusions from realities refutes the *esse est percipi* doctrine. Second, the very notion of a phenomenon or appearance is unintelligible apart from its dual, that of objective fact. (Thus

it makes sense to estimate real sizes and relate them to apparent sizes only because we can distinguish them.) Third, every indirect observation, particularly if quantitative, employs hypotheses (such as law statements) that go beyond phenomena. Fourth, whereas some facts are observer-invariant, i.e. the same for all knowing subjects, others are not—and it is a task of scientific research to find out which are what. Fifth, the present can be explained only by assuming—as cosmology, geology, evolutionary biology and other sciences do—that there were real things before animals capable of perceiving them evolved. In short, the collection of all possible phenomena is only a (smallish) subset of the collection of all possible facts.

Yet, however insignificant appearances may be from an ontological point of view, they occupy a central position in epistemology. In fact, there is no way of gaining some deep knowledge about reality except by combining phenomena with hypotheses and processing both by reasoning. Not that, as James held (1890, II, p. 7), "Sensations are the stable rock, the terminus a quo and the terminus ad quem of thought"—for perception is fallible and has a very restricted scope. Phenomena are central to knowledge for different and various reasons. First because they *pose problems*—e.g. "Why do we perceive X?", "Why cannot we perceive X?", "Is X real or illusory?", and "Why does X appear to us now as Y, now as Z?" Second, because phenomena, when under control, *test* factual hypotheses—as in "If X is the case, then we must be able to observe Y. Now, we failed to observe Y. Therefore X cannot be the case." As James (1890, II, p. 301) put it, "What science means by 'verification' is no more than this, that no object of conception shall be believed which sooner or later has not some permanent and vivid object of sensation for its *term*. [. . .] Sensible objects are thus either our realities or the tests of our realities. Conceived objects must show sensible effects or else be disbelieved. [. . .] Strange mutual dependence this, in which the appearance needs the reality in order to exist, but the reality needs the appearance in order to be known!"

Since phenomena constitute only a very special kind of fact, real existence and the totality of real existents (i.e. reality) cannot be defined in terms of phenomena. In particular it is incorrect to define a real thing as a permanent possibility of being perceived (Mill, 1843, Bk. III, Ch. 24, Section 1). On the contrary, we must define appearance, or the totality of phenomena, as the collection of all (actual or possible) perceptual processes in all animals past, present and future. (We may also specify and speak of human appearance, blue jay appearance, sardine appearance, etc.) But,

since we test existence hypotheses with the help of phenomena, the latter do provide a criterion of real existence. We propose the following *criterion* (not definition) of objective existence:

X exists only if the following conditions are met jointly:

(i) there are processes in X that produce, directly, or indirectly (i.e. via some other item), some perceptible effect Y;

(ii) this effect is observed by any competent observer in similar circumstances;

(iii) the link between X and Y can be rationally justified (preferably in terms of law statements).

Clause (i) includes the most commonly accepted reality criterion, namely "Being is acting (or having efficacy)". Since the universe as a whole cannot act upon anything, yet is real, this criterion, unlike ours, leaves out the universe. Clause (ii) is the condition of publicity or sociality of objective knowledge: an existence hypothesis proposed by a single individual, whether crackpot or genius, will not be accepted unless independently confirmed. And clause (iii) is the rationality condition: a dogmatic existence statement is unacceptable, for we must understand, at least in outline, the link between the phenomenon and the purported objective fact.

The above existence criterion is peculiar to *critical* (as opposed to naive) *realism*. This class of epistemological doctrines is contrasted in Table 4.1 to alternative views.

TABLE 4.1
Views on the relation of existence to perception.

Thesis	*Antithesis*
1. All perceptibles are real. (*Naive realism*)	1′. Some perceptibles are not (objectively) real.
2. All existents are perceptible. (*Naive phenomenalism*)	2′. Some existents are imperceptible.
3. All and only perceptibles are real. (*Phenomenalism*)	3′. Some existents are imperceptible or some perceptibles are not (objectively) real.
4. No existents are perceptible. (*Buddhism*)	4′. Some existents are perceptible.
5. All imperceptibles are real. (*Platonism*)	5′. Some (objects) are neither real nor perceptible.

4. Concluding remarks

Only philosophers have questioned the principle that perception is necessary for knowing reality, and only philosophers have claimed that it is sufficient. Everyone else agrees that perception is necessary though insufficient: that it does yield knowledge (perceptual knowledge), though low grade and fallible. We all know that one and the same external object may appear small to an adult and big to a child, warm to an observer coming from cold outdoors and cold to another staying indoors. We all realize that we are not good at estimating directly vertical distances, very small or very large weights, and mental or social states. And many of us know also that by far the most numerous properties calculated or measured in factual science are just imperceptible. But we also know that any such properties, to be known, must be somehow related to sensible ones—via hypotheses, instruments, or both. True, perception is often deceptive: Figure 4.6. However, we can often discover and correct such perceptual errors, e.g. by checking with other senses, or other people, or by scientific analysis.

What holds for perception holds also for action and, indeed, for all types of experience: experience is no *solid* ground of knowledge but it is (metaphorically speaking) one of the two inputs to our cognitive apparatus—the other being reason. (Strictly speaking all experience is either past or future, hence hardly a solid basis for anything.) Strictly experiential judgments, of the forms 'I feel *X*', 'I do *Y*', 'I feel *X* when I do *Y*', 'I do *Y* when I feel *X*', and the like, are not objective because they report on a changing subject-object relation rather than on a subject-free object. Objective statements do not include reference to the subject. Such objectivity can be attained in either of two ways: really or illusorily. The

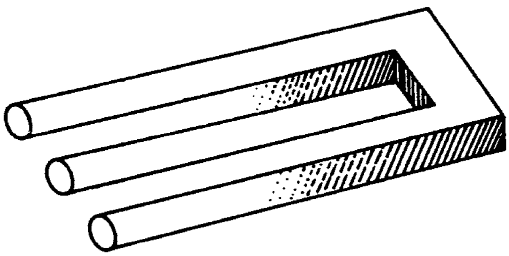


Fig. 4.6. The Devil's Trident: physically impossible. An illusion produced by trickery.

former by thinking up hypotheses about facts in themselves, i.e. independent of the way they are perceived, and putting such hypotheses to the test. This is the way of science and technology. The illusory way of attaining objectivity is the linguistic trick of substituting the impersonal 'It feels *X*', '*Y* is done', etc., for the personal 'I feel *X*', 'I do *Y*', etc.

Science and technology strive for objectivity, i.e. for knowledge of things in themselves. In other words science and technology discover or make facts but do not invent or construct them: what they do invent or construct is concepts, hypotheses, theories, methods, plans, and artifacts. This is not the view of subjectivistic philosophies, whether extreme like Berkeley's or moderate like Kant's. According to these philosophies there are no objective facts or, if there are any, they are unknowable: all facts would be man-made.

There are two kinds of subjectivism (or radical constructivism): individualist and collectivist. According to the former the individual knowing subject constructs the (his) world; according to collectivist constructivism, the world is an outcome of a collective activity. Individualist subjectivism is well known from classical philosophy, but it has hardly any followers among contemporary philosophers. Paradoxically it has some among theoretical physicists, particularly those who, following Bohr (1935), are wedded to the operationalist interpretation of quantum mechanics and state that nothing is an event ("phenomenon") unless it is observed by someone. Thus we are told that "We know now that the moon is demonstrably not there when nobody looks" (Mermin, 1981). (For a criticism of the subjectivistic misreading of quantum mechanics see Bunge (1973) and Vol. 6.) The following paradox demolishes this kind of subjectivism. If a subjectivist claims that nothing exists unless it is observed, then he must own that his own brain cells do not exist as long as their activity is not individually recorded. But this involves inserting an electrode in each of them, which, if technically feasible, would kill him. So, if the subjectivist wants to make sure that he is alive, he must allow a neuroscientist to kill him. It is better to give up subjectivism and stay alive.

Collectivist subjectivism is newer and less well known although it is rapidly gaining in influence. It was first formulated by Fleck (1935), adopted in a less extreme form by Kuhn (1962), and literally by a number of anthropologists and sociologists of science (e.g. Berger and Luckmann, 1966). Thus Latour and Woolgar (1979) claim to describe how Nobel laureate Roger Guillemin "constructed rather than discovered the fact of

the thyrotropin-releasing factor". (Moreover he would have discovered such fact not as an outcome of observation, experimentation and hypothesis, but as a result of monitoring a number of "inscription devices" leading ultimately to the final inscription process called 'writing a paper', for which credit can be obtained. They do not explain why the whole process could not have been entrusted to a computer.) The only advantage of this type of constructivism is that it is self-destructive.

Empiricism held that perception is our point of departure, everything else being built upon it. (According to objectivist empiricism, such as Locke's, what we add is conception. On the other hand radical empiricism, such as Mach's, and the later James', claims that that world itself is made up of sensations.) Psychology has shown that there is no perceptual

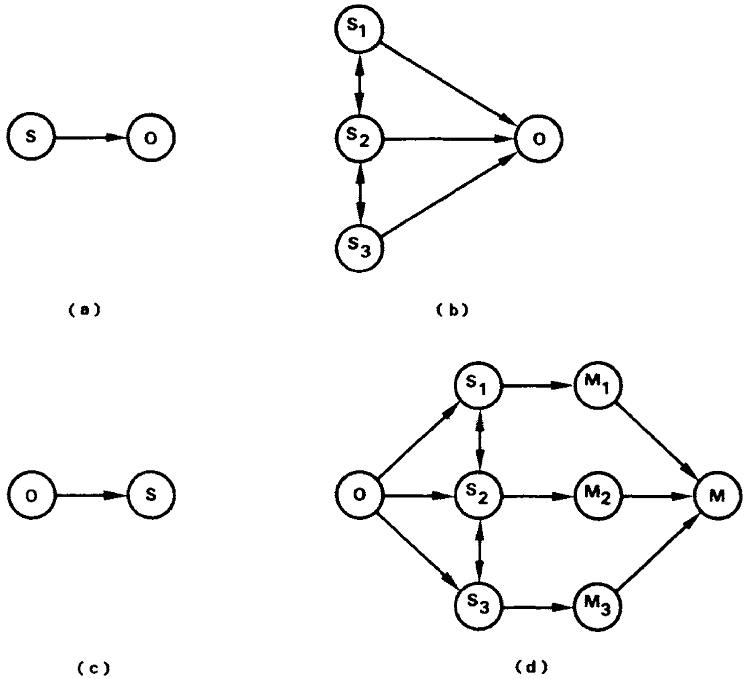


Fig. 4.7. Four views of the subject-object relationship. (*a*) *Individualist subjectivism*: the subject constructs the object, which is either a complex of perceptions (phenomenalism) or a theory (rationalism). (*b*) *Collectivist subjectivism*: the object is constructed by a community of subjects. (*c*) *Individualist realism*: the subject copies or mirrors the object. (*d*) *Social critical realism*: inquirers interacting in a community construct alternative models M_1, M_2, etc. of the object, which are tested until a model M is adopted by most of the members of the community.

"given" to be formulated as a set of indubitable basic sentences. We produce our own perceptions which are colored by our preconceptions, expectations, and social circumstances. All knowledge, whether perceptual or conceptual, is constructed—though not always deliberately.

What we are given is our external world and our genome. We do not have the burden of constructing the whole of reality: we are born into a self-existing world, which we manage to enrich in some respects and to mess up in others. And we are born with certain cognitive abilities that the natural and social environment can enhance or dim—in short, educate. Our perceptual limitations drive us not to constructing reality but to conceptualizing it. And the incompleteness and uncertainty of both individual perception and individual conception advise us to supplement them with, and check them against, those of similar inquirers. The result is that socially related knowing and acting subjects build fallible perceptual and conceptual representations of reality. This is the gist of our epistemology, which is a variety of realism, namely *biological* (as opposed to mentalistic), *social* (as opposed to individualistic), and *critical* (as opposed to naive and dogmatic). See Figure 4.7.

CHAPTER 5

CONCEIVING

We now come to ideas proper, in particular concepts and judgments. Other types of ideas, such as reasonings, problems, and rules, will be treated in subsequent chapters. Although some ideas originate in perception they all overreach it. Sit quietly in a sound-proof room, close your eyes, and think of any intellectual problem. You will be thinking up concepts, most likely not isolated concepts but entire systems of them.

We shall start by examining elementary ideas, i.e. concepts, such as those of book and equality. There are two different concepts of a concept: the psychological and the philosophical one. In psychology a concept is a mental (or brain) process; in philosophy it is a unit construct, or the smallest component of a proposition. In either case concepts are the building blocks of all other ideas. We relate the two concepts when we say that we are thinking of a given concept. Thought is a concrete process, hence one that takes time; on the other hand its object is abstract, in the sense that we detach it from such process and feign that it exists by itself as a timeless object. In this chapter we shall deal with the two notions of a concept, as well as with judgments and their philosophical counterpart, namely propositions.

1. Concept

1.1. *From Percept to Concept*

When seeing or imagining a particular animal we say that we form a *percept* of it—or, more simply, that we *perceive* it. Such a percept, to have any cognitive value at all, must be in some respects analogous to the object perceived or imagined. On the other hand when thinking of humans in general—e.g. that they can conceive anything—we form *concepts* of them, e.g. those of *Homo sapiens, Homo faber, zoon politikon*, etc. Although a concept of this kind derives partly from our perceptions and memories of individual human beings, it is not identical with them. To begin with it is general, and so can bear no resemblance to any individual. A fortiori the so-called abstract concepts, such as those of transitivity, entropy, species,

mind, or justice, are not similar to their referents even when they have them.

In other words, the perceptual representation of a thing is a (partial) function of its (apparent) traits into neural events, that preserves some of the relations among such traits (Ch. 4, Section 3.1). On the other hand a concept, or conceptual representation, of the same thing is no such map: it bears no resemblance to its referents, if any. Think of a theory of the electromagnetic field or of the competitive economy, each compressed into a system of equations. This does not mean that, unlike percepts, concepts are in an immaterial mind, let alone an impersonal Mind, rather than in some concrete brain. Concepts are brain processes (or collections of such) but they do not involve any neurosensors even though some of them do originate in perception.

In the above we have tacitly distinguished two kinds of concept: those which originate in perception and those which do not. We may call them *empirical* and *transempirical* (or transcendental) respectively. Idealists deny the existence of the former or render it dependent upon that of the latter. On the other hand empiricists and vulgar materialists (nominalists) deny the existence of transempirical concepts, and hold consequently that the words designating them are just meaningless sounds. Thus whenever Hume (1739) set out to evaluate an idea he asked for its psychological pedigree: *From what (sense) impression cou'd this idea be deriv'd*? Since he could not find the sensory roots of the metaphysical ideas of substance, cause, and self, he dismissed them. Likewise Kant (1787) professed to reject all the concepts of pure reason, particularly if they purported to refer to things in themselves: he too wished to stick to phenomena although he never succeeded. Needless to say, both Hume's and Kant's criticism of transempirical (in particular metaphysical) concepts holds only provided one accepts the empiricist and anthropocentric dogma that only that exists which can make some sense impression on us. (Recall Ch. 4, Section 3.2.)

If Kant and Hume had known the mathematics and physics of their time, they would have rejected them with the same vehemence with which they rejected metaphysics, for those sciences contained key transempirical concepts such as those of function, continuity, and infinity, speed of light, action at a distance, and mass. And they would have disapproved even more strongly of contemporary science, which involves a far more abstract mathematics, as well as factual concepts far removed from experience, such as those of field, electron, and gene. In particular, as Einstein noted, the concepts of theoretical physics are independent of sense experience and are not abstracted from it. "The relation [between concept and percept] is not

analogous to that of soup and beef but rather of wardrobe number of overcoat" (1936, p. 353). Consequently "There is no inductive method which could lead to the fundamental concepts of physics" (*op. cit.*, p. 365).

To be sure the concepts of the empirical kind "derive" from percepts. However, they are not mere distillates or summaries of experience. The coining or learning of a new concept of either kind is a creative process: it consists in the emergence of something new that was not in perception, let alone in the external world. (From a neurophysiological point of view it consists in the emergence of a new plastic neuronal system: recall Ch. 1, Section 1.2. Such a process occurs in one of the higher brain centers. This explains why any empirical concept can be aroused by any one of the senses.) We may think of such a process as a map composition, or map of maps: from stimulus objects to percepts and from these to empirical concepts. See Figure 5.1.

The transempirical concepts do not originate in perception, i.e. they cannot be learned from experience but must be acquired by reflection. However, they are not necessarily isolated from all empirical concepts. For one thing some transempirical concepts may be suggested by empirical ones. Thus the empirical ideas of five fingers, five shells, and five children may have suggested forming the transempirical concept of number five. Conversely, some transempirical concepts may suggest empirical concepts and these in turn may guide some experiences. For example, thinking of low temperatures may suggest the idea or the image of a cold drink, which may in turn elicit the idea of paying a visit to the fridge.

In sum, although we emphasize the differences between percept and concept, and between empirical and transempirical concepts, we also assert that the items of these three kinds are not isolated but compose three-tiered *dynamical cognitive networks*. Each of them is related to some others, and each of them can modify some others. This view subsumes both empiricism and idealism, for it admits the creativity of both experience and the

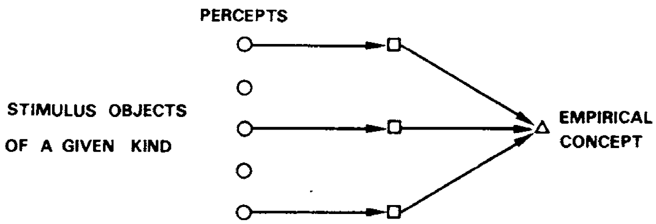


Fig. 5.1. The formation of empirical concepts as map composition.

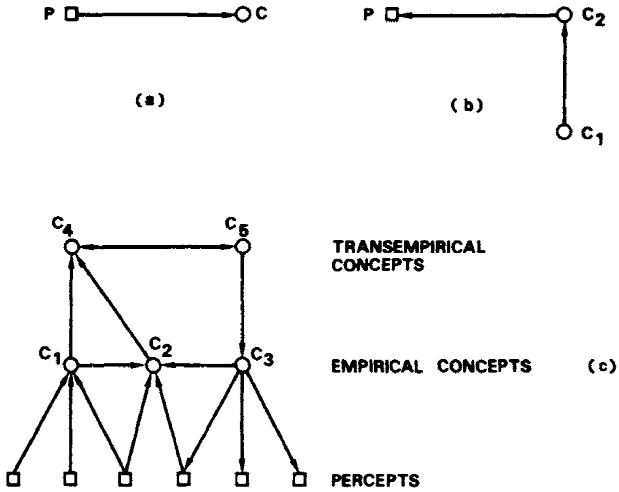


Fig. 5.2. Three views on the relations among percepts and concepts. (*a*) *Empiricism*: Every concept originates in a percept. (*b*) *Idealism*: Every concept originates other concepts, which in turn make some percepts possible. (*c*) *Realism*: Whereas some concepts originate in percepts, others do not; besides, some concepts guide perception and all concepts generate further concepts.

mind—or rather the brain. (See Figure 5.2.) But it transcends the borders of both classical epistemologies in that it suggests looking for the neuronal systems that do the perceiving, the empirical conceiving, and the transempirical conceiving as well as for the mutual actions among such systems.

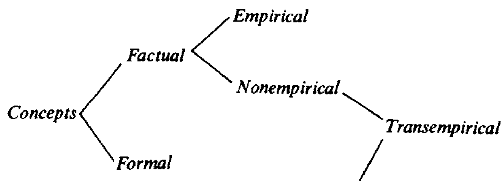
(Note that we are presupposing that perceiving and conceiving are activities of multineuron systems. The idea that there is one neuron for every concept lacks experimental support and is inconsistent with the existence of about 10^{11} neurons in the human brain. Indeed even a supergenius capable of learning one concept every minute would have accumulated, after a long lifetime, only about 50 million concepts, a tiny fraction of the total number of plastic neurons in the human brain: the vast majority would be idle.)

Among the transempirical concepts we distinguish the formal ones, i.e. those occurring in logic, mathematics, and semantics. These concepts are called 'formal' because they do not refer to any facts, whether objective or subjective (experiential). Thus the logical concept of conjunction concerns predicates and propositions, the mathematical concept of limit concerns functions and sequences, and the semantical concept of

reference concerns predicates and propositions. That such concepts have no factual reference can only be proved with the help of some theory of reference. (Cf. Vol. 1, Ch. 3.) That, in spite of such lack of factual reference, they can be "applied" in factual science and technology, is another matter to be discussed in Vol. 6.

The admission of formal concepts was a major sacrifice for contemporary empiricism (or logical positivism): it had to be made in order to use logic to analyze and organize empirical knowledge. On the other hand the materialist suffers no setback in admitting formal concepts: all he needs is the hypothesis that all concepts, even the formal ones, are brain processes or collections of such (Bunge, 1981a).

According to logical positivism all facts are experiential, so that all concepts with a factual reference must have an empirical content or meaning, i.e. must be empirical concepts. Hence according to that philosophy the factual/formal dichotomy coincides with the empirical/formal one. This classing is wrong because it is not exhaustive. Indeed there are nonexperiential facts, such as atomic collisions and species extinctions, and we have formed the corresponding transempirical concepts. In other words, whereas some factual concepts are empirical (e.g. "hot") others are transempirical (e.g. "temperature"). This is a firm result of a semantical analysis of scientific concepts (Bunge, 1973, 1974a, 1974b). The concept classification that accounts for the (conceptual) existence of nonempirical factual concepts is this:



Are there any laws of the development of concepts, or rather of our conceptual abilities? We know next to nothing about the conceptual development of children before the first year of age, and this mainly due to their very poor motor and communication abilities. However, neuroanatomical studies show that the cortex of the infant brain is poorly organized and therefore unlikely to perform any conceptual feats: it has small neurons, incompletely myelinated axons, a small dendritic arborization, and few synaptic connections. Older children have been investigated by Piaget and his school for over half a century. These studies have found that there are indeed definite stages in development that cannot

be skipped. In particular they have found that the young child has hardly any general concepts: he or she acquires them in the course of development—the pace of which is presumably accelerated by formal education. Once the child has acquired some general concepts—such as that of causation—and some general propositions—such as the principle of conservation of matter—he can use them to construct further concepts as well as to face experience in novel ways.

As for conceptual evolution, we know even less. However, it is likely that all hominids, and perhaps many other higher vertebrates as well, have formed certain general concepts, such as those of existence, presence, place, time, notion, lightness, hotness, hardness, nearness, usefulness, and friendship. A methodical study *à la* Piaget of the conceptual abilities of apes, dolphins, ravens and other animals usually rated as intelligent, is overdue. Also, it would be interesting to investigate the tacit knowledge that must have gone into the fashioning of artifacts by primitive and ancient man. We know little about this in part because of the epistemological prejudice that every concept is preceded by some percept or some action. We know much more about the history of ideas. One thing we have learned from it is that, at least from the beginning of civilization, it is just not true that every complex idea has been preceded by simpler ideas. Whereas much progress consists in complication, some of it involves simplification. Besides, some very basic ideas, such as those of set and concatenation, were introduced long after far more complicated ideas—e.g. those of the infinitesimal calculus—had been invented. In short, there are laws of conceptual development but not of conceptual evolution.

1.2. *Concept Formation*

Let us now examine a handful of very basic conceptual operations. They are so basic that we usually take them for granted and therefore do not pause to analyze them. After all, it behoves philosophers to puzzle what others take for obvious.

The most basic relation among any two items, whether factual or conceptual, is that of difference. The perception or conception of difference is called *discrimination* (or *analysis*). Its dual, the failure to differentiate, is of course confusion. Our perceptual apparatus has a very restricted discriminatory (or analytic or resolving) power, but we can enhance it with the help of observation instruments. Likewise we can sharpen conceptual discrimination or analysis with the help of logic and mathematics.

We must guard against the temptation of detaching whatever is distinguishable. For one thing, properties and changes thereof (i.e. events) are not detachable from the concrete things that possess the properties or undergo the changes. (See Vol. 3, Ch. 2.) For another, every thing is a system or a component of some system, and every system but the entire world interacts with some other system. (See Vol. 4, Ch. 6.) The components of a system, though distinct and therefore distinguishable (at least in principle), are not separable. (Consequently it is impossible to specify the state of one of them without reference to the state of the others. I.e. the state space of the whole does not equal the union of the state spaces of the components.)

All of the mistakes alluded to above have been erected into philosophical doctrines. Thus (philosophical) *atomism* conceives the world as a heap (not a system) of detached or separate units (atoms): it has no use for systems possessing (emergent) properties not possessed by their components. Its dual, namely *holism*, prefers confusion: it proclaims that whatever cannot be detached is one. (See Vol. 4, Ch. 1.) Holists abhor analysis believing that it destroys wholeness, while actually it is the only way to understand wholes. Finally *dialectics* declares differents to be opposites, while at the same time proclaiming the unity of opposites. Since it does not analyze the concepts of opposition and of unity, dialectics is not easy to understand, hence to evaluate. (See Bunge (1981a) for a criticism.)

Discriminating between two objects is a modest cognitive achievement—even though, as in astronomy, it may call for an enormous cognitive investment. The next step is to find out what the distinct objects differ in. And this involves taking cognizance of some of their properties. Such cognitive operation is no mere perception: it is a conceptual operation, for it consists in attributing properties. Indeed, in order to be able to attribute a property to an object we must form some concept of such property. This concept may precede the act of attribution or it may be improvised during it, but in any case it must be formed in order for predication to occur. A concept of a property is called an *attribute* or *predicate*. (There are of course unary, binary, and in general n-ary predicates. Thus the predicate *between* is ternary, and so is any function of two variables.)

Idealists need not adopt our distinction between properties and predicates, but realists must, and this for the following reasons. First, a property of a thing cannot be detached from the latter: there are no (substantial) properties without things, and no things without properties

(Vol. 3, Ch. 2). On the other hand a predicate, if attributed (truthfully or falsely) to a concrete thing, is a conceptual representation of a thing's property. Second, attribution is a psychological process, hence one studied by psychologists, whereas all sciences are supposed to study properties objectively possessed by objects. Third, one and the same property can often be conceptualized (represented conceptually) by more than one predicate. Thus on a first level of analysis heaviness is representable as a unary predicate ("body b is heavy"); on a second level as a binary predicate ("body b in gravitational field c is heavy"); on a third level as a ternary predicate ("the heaviness of body b in gravitational field c equals d"), and so on. (In general, predicate formation can be construed as a function from properties to sets of predicates.) Fourth, whereas for every attribute there is another attribute equal to the negation of the first, things have only "positive" properties. A thing either possesses P or does not possess P, but it cannot "possess" not-P: negation is *de dicto* not *de re*. (Barnacles do not think, but this is not to say that they exert the function of not thinking. Negation affects the proposition "Barnacles think", not the property of thinking.) Fifth, what holds for negation holds also for disjunction: there are no disjunctive properties although there are disjunctive predicates. There is no such thing as the property of being alive *or* dead, although the predicate "is alive or dead" is perfectly respectable, and moreover tautological. In short, whereas the set of predicates of a given rank and a given reference class forms a Boolean algebra, the corresponding set of properties does not. (See Vol. 3, Ch. 2.)

There are two different ways of forming the concept of a property, i.e. of construing a predicate: intensionally or extensionally. In the former construal the predicate is a function from individual objects to propositions (Vol. 1, Ch. 1, Section 1.2). For example, "Metabolizes" assigns to organism b the proposition "b metabolizes", and "Interacts" assigns to the ordered pair of things a and b the proposition "a interacts with b". The truth value of such propositions is immaterial for the construction of the predicate. However, once we have found out such truth values we can form the extension of the predicate, i.e. the collection of individuals (or pairs, or n-tuples) for which the predicate holds. And then we can ascertain whether a given individual (or n-tuple) belongs to the extension of a certain predicate. In other words, the ordinary language expression 'b is F' can be construed either as 'Fb', where F is a function, or as 'b is a member of the extension of F'.

Ordinarily the extensional construal presupposes the intensional one, for

we must know what predicate we are talking about (and what property it conceptualizes) before inquiring into its extension. However, both construals are equivalent, and mathematicians use them interchangeably. (For example, the axiom of mathematical induction, occurring in arithmetic, can be formulated in either of the following ways: (i) If a property holds for zero and, when holding for an arbitrary number, it also holds for its successor, then it holds for every natural number: (ii) If a set of natural numbers contains zero, and contains also the successor of every natural number it contains, then it contains all natural numbers.)

Nominalists trust only individuals and mistrust concepts and therefore predicates. To them every property must be understood as the collection of individuals possessing it: they conflate predicates with their extensions. This confusion is open to several objections. The most damaging ones are as follows. First, let P_1 and P_2 be two properties of entities of a certain kind K, e.g. the mass and momentum of bodies. Since all K's possess both P_1 and P_2, according to nominalism P_1 is identical to P_2—which contradicts the hypothesis that P_1 and P_2 are different. Besides incurring contradiction, extensionalists and nominalists are confronted with a practical difficulty: they must know all the values of a given function in order to be able to speak of the function. But in practice this is seldom the case: we are not given complete tables but differential equations, recursion formulas, or some other condition defining the function. Moreover we must know such defining conditions in order to be able to compute (some) values of the function. (In addition we need some special technique, such as series expansion, or iteration, or numerical interpolation.) In short, we cannot sacrifice science and mathematics to philosophical prejudice.

When forming concepts we should avoid the dual sins of reification and ideafication. *Reification* is of course the incorrect conception of properties, relations, or even concepts, as entities having an autonomous existence. Classical examples are the idea that sickness is an entity that the patient carries and may pass on to someone else, and the medieval conception of darkness as an entity (*tenebras*). More recent examples are the existentialist notion of nothingness as a thing, and the structuralist idea that structures precede the corresponding structured things. The dual procedure is *ideafication*, or the conceiving of material things or processes as self-existing ideas. Classical examples are the idealist conception of ideas as detached from thinking brains, and of culture as a system of ideas, values, norms, and behavior patterns. To be sure it is permissible, nay sometimes indispensable, to *feign* that mental processes are detached from brains, in

order to be able to focus on certain common features of concepts and disregard everything else, in particular the thinkers' circumstances. This is merely an instance of what may be called *methodological abstraction*. But we should keep in mind (or rather brain) that such conception of concepts in themselves is a fiction.

In studying properties we shall do well to distinguish three kinds of them: (*a*) *individual* or possessed by individuals (e.g. the mass of a molecule, the demand for a certain commodity by a certain individual consumer); (*b*) *aggregate* or possessed by aggregates (e.g. the total mass of a conglomerate of bodies, the total demand for a certain merchandise by a society); (*c*) *emergent*, or possessed by a system but not by its components (e.g. the stability of a molecule, the type of an economy). Failure to draw these distinctions results in absurds such as "the average molecule" or "the average consumer" (instead of "the average of property *P* over an aggregate of molecules—or consumers"). Two cautions are in order. The first is that a number of properties that can be treated as aggregative (hence additive) to a first approximation are actually systemic. This is the case with mass and entropy, as well as with aggregate demand and supply in economics. (Consider: if two blue collar families are friends and neighbors, they can pool their resources to buy one car, whereas separately neither of them might afford to.) Second, a number of properties that seem to be possessed by individuals are actually possessed by components of a system and vanish if the system breaks down. (For example, the property of being a faithful spouse vanishes upon divorce. Likewise some of the properties of producers and consumers depend on the type of economy.)

There are essentially three ways of studying properties: (*a*) to study *particular* properties, such as the surface tension of a given liquid or the feeding habits of animals of a given species in a given environment; (*b*) to study the *distribution* of a given property in a population, e.g. the distribution of fertility in an animal population, or of wealth in a human society; (*c*) to study the *relations between two or more properties* of a thing, e.g. the correlation between unemployment and crime. All three are legitimate concerns, so neither should exclude the other, and all three should be carried out both theoretically and empirically.

Perceiving or conceiving commonalities is no less important, for attaining knowledge, than realizing differences. Two items are said to be *equivalent* (or *equal*) in some respect, i.e. with respect to a given property, just in case they share this property. This definition entails that there are as many equivalence relations as there are properties. It also entails that every

equivalence relation is reflexive, symmetrical, and transitive. I.e., for any objects a, b and c in the given set: $a \sim a$ (reflexivity), if $a \sim b$ then $b \sim a$ (symmetry), and if $a \sim b$ and $b \sim c$, then $a \sim c$ (transitivity).

If the transitivity condition is relaxed we obtain the weaker relation of *similarity* or *resemblance*. If a is similar to b, and b similar to c, it may or may not be the case that a resembles c. That is, a and b may share some properties which are not those that b and c share: see Figure 5.3. A qualitative measure of similarity is the intersection of the sets of properties of the similar objects. That is, calling $P(a)$ and $P(b)$ the sets of properties of a and b respectively, we stipulate that the similarity $s(a, b)$ between a and b equals $P(a) \bigcap P(b)$. By counting the number of shared properties we obtain a quantitative measure. Our qualitative measure allows us to introduce two relations of similarity comparison:

$$a \textit{ is more similar to } b \textit{ than to } c =_{df} s(a, b) \supseteq s(a, c),$$

$$a \textit{ is more similar to } b \textit{ than } c \textit{ to } d =_{df} s(a, b) \supseteq s(c, d).$$

If two items are distinct (i.e. really two instead of one), and yet equivalent in some respect, then they can be put into a single class. (Not so if they are merely similar.) Note the expression 'can be put' instead of 'exist'. Differences among factual items are objective: they exist whether or not a given subject knows it. On the other hand grouping is a conceptual operation: classes are concepts, not real (concrete, material) entities. But of course whereas some groupings are arbitrary or artificial others are natural or objective. Thus by putting together all the persons called John we obtain an artificial class, whereas by grouping all the persons sharing some common ancestors we obtain a natural class, i.e. a collection whose members are objectively related. (For the notion of natural class see Vol. 3, Ch. 3.)

The operation of grouping or collecting is a basic animal behavior: all animals gather things of some kind or other, and the higher vertebrates are also capable of collecting ideas, i.e. of grouping images, concepts, propositions, and the like. Material things can be put together spatially or

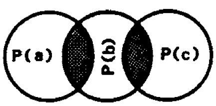


Fig. 5.3. The properties shared by a and b are not the same as those b and c have in common. Hence, although a is similar to b, and b similar to c, a does not resemble c.

temporally; and all objects, regardless of their nature, can be put together in thought. In the former case the outcome is a material thing: a conglomerate or a system; in the second case the outcome is a concept: a set. And in either case the outcome is assumed to be independent of the order in which the items have been gathered: i.e. the result is either an unordered aggregate or an unordered set—or perhaps the matter of ordering has been disregarded on purpose. We are concerned here with collecting items of any nature whatever, whether factual or conceptual, to form conceptual collections, i.e. classes. Whereas some classes are realistic (not arbitrary), no class is real. We shall come back to this subject in relation to biological taxonomy (Vol. 6).

(The operation of collecting or grouping can be defined as follows. Consider three items: $0, a$, and b, forming the following binary compounds or concatenates: $00, 0a, a0, 0b, b0, aa, bb, ab$, and ba. Assume that $00 = 0$, $0a = a0 = a$. $0b = b0 = b$, $aa = a$, $bb = b$, and $ab = ba$. In other words, 0 is the null object, and concatenation is idempotent and commutative. We now define the *collecting* or *bracketing* operation $\{\ \}$ by the following assignments:

$$0 \mapsto \{\ \};$$

$$a \mapsto \{a\}, \qquad b \mapsto \{b\};$$

$$ab \mapsto \{a, b\}.$$

From the assumed properties of the concatenation it follows that

$$aa \mapsto \{a\}, \quad bb \mapsto \{b\}, \quad ba \mapsto \{a, b\}.$$

In the above, a and b may be any objects, whether material or conceptual. On the other hand 0, the null object, can only be conceptual: nothingness has no real existence. Finally, we define $\varnothing =_{df} \{\ \}$.)

When collecting real (material) things we must recall that they are changeable, to the point that whereas some of them may have already vanished others may not yet have come into being. (Think of mankind, or the human species.) We must therefore distinguish between a collection that varies in time and a set, or collection regarded *sub specie aeternitatis*. The membership of the former is variable, whereas that of the latter is constant: all the elements in the latter are treated as if they were timeless entities. The difference between the two collecting operations can be elucidated with the help of the notion of time, which is of course a scientific and ontological concept, not one in pure mathematics. In fact call F the

attribute of interest and write 'Fxt' to indicate that x has property F at time t. Then the *collection of all F's at time t* will be

$$S_t = \{x \mid Fxt\}.$$

(Instead of augmenting the n-arity of predicates we can also use the operation of bracketing at a given instant or over a given time interval, or $\{\ \}_t$ for short. This operation can be defined by adding t to the individuals $0, a$, and b used to define the timeless bracketing operation $\{\ \}$, and requiring that it obey the following conditions:

$$0t \mapsto \{\ \}_t;$$

$$at \mapsto \{a\}_t, \qquad bt \mapsto \{b\}_t;$$

$$abt \mapsto \{a, b\}_t.$$

The expression '$\{a, b\}_t$' is read 'the set formed by a and b at (or during) t'.)

By taking the union of all the variable collections of F's for all time we get the (timeless) *set* of F's, or the *extension* of the predicate F. (In physics many predicates are frame dependent, so it may be necessary to add a subscript naming the particular reference frame relative to which the collecting is done.)

Just as we can analyze individual things and individual ideas, so we can analyze collections of either. The most basic mode of analysis of a collection is its *partition* into homogeneous subcollections. For example, a human population can be analyzed into collections of individuals of the same sex, the same age, or the same number of schooling years, etc. The key to such partition is the concept of sameness in some respect, i.e. of equivalence, examined a while ago. One says that an equivalence relation *induces the partition* of a collection into a family of *equivalence classes*, and that all the members of each equivalence class are equivalent (or equal) in the given respect. The operation resulting in the formation of such family is called the *quotient* of the given set S by the given equivalence relation $\sim$ (written '$S/\sim$'). For example, the quotient of the set of integers by the relation of equal parity is the family composed by the set of even numbers and the set of odd numbers. (I.e. $\mathbb{Z} = \{E, \bar{E}\}$, where $E = \{x \in \mathbb{Z} \mid (\exists y)(x = 2y)\}$, and $\bar{E} = \{u \in \mathbb{Z} \mid (\exists v)(u = 2v + 1)\}$.) We shall use this notion when studying the principles of classification (Ch. 9 Section 1.1).

It has been noted that, whereas some people excel at analyzing, others are better at synthesizing: few are at the same time good splitters and good

lumpers. Yet both abilities are needed, for whereas some lines of inquiry start by distinguishing or resolving, others begin by fusing or lumping, and all research projects include these two moments. Indeed, analysis does not go far without interrelation, and in turn interrelation may result in the recognition of wholes. So, splitters as well as lumpers discharge a useful function in any inquiring community—provided each of them does not enshrine his own ability as a philosophy (atomism and holism respectively) and proclaims the superiority of it.

So far we have encountered concepts of three basic types: individuals, collections of such, and functions. In the present stage of the evolution of ideas every concept is either an individual concept (e.g. "Earth"), a collection (e.g. that of all solar planets), a function (e.g. "not"), or some construction out of concepts of such basic types. For example, an ordered n-tuple may (but need not) be construed as a certain set of sets. (E.g. $\langle a, b\rangle = \{\{a\}, \{a, b\}\}$.) A vector may be analyzed as an n-tuple of real numbers and also as a unit component of a vector space.

The general concept of an n-tuple and the more special one of a vector are formal concepts. However, they can be applied to real life situations. Consider a set of commodities (goods or services). Call q_i the quantity of the ith commodity, measured in appropriate units, and p_i the corresponding unit price in dollars. Then the value of the amount q_i of commodity i is by definition the product $p_i q_i$. A collection or bundle of commodities q_1, $q_2, \ldots, q_n$ can be conceptualized as an ordered n-tuple. Prices are parallel. The price structure or value of a bundle of commodities is then defined as the product $v(q) = pq$ of the vectors $p = \langle p_1, p_2, \ldots, p_n\rangle$ and $q = \langle q_1, q_2, \ldots, q_n\rangle$. This notion allows one to compare the values of different commodity bundles. In fact we stipulate that $v(q) \geq v(q')$ if and only if $pq \geq p'q'$. In this manner we can conceptualize a whole bundle of commodities as a single concept, and reduce the comparison of two bundles of commodities to comparing two numbers. Moreover all of the concepts involved in the above process are exact.

The general concept of an n-tuple is abstract, that of an n-tuple of real numbers (or vector) is less so, and that of a commodity bundle is not at all abstract. This kind of abstraction may be called *semantical* because it is the dual of semantic interpretation. In fact we get the notion of a vector by interpreting the components of an n-tuple as real numbers, and the notion of a commodity bundle by interpreting in turn those real numbers as quantities of commodities. In other words the factual concept (of commodity bundle) results from the factual interpretation of a mathemati-

cal interpretation of an abstract concept. (For the composition of the two interpretation maps see Vol. 2, Ch. 6.) Semantical abstraction is, in short, the dual or inverse of interpretation.

Another concept of abstraction may be called *methodological*, as it consists in deliberately disregarding certain properties. It is identical with that of equivalence. Indeed, if in comparing two objects we abstract from (disregard) all their properties except those they share, we build an equivalence relation allowing us to partition the given set into equivalence classes. (Recall what was said about equivalence a while ago.) An instance of this notion of abstraction is the formation of the concept of relation apart from the relata. Thus take all the possible statements of the form "x lies to the right of y", and abstract from the relata, i.e. consider the relation of lying to the right: this concept is methodologically abstract. Clearly, every time we generalize and every time we build a model of a real thing we disregard certain properties and thus perform a methodological abstraction.

A third concept of abstraction is *epistemological*, for it consists in remoteness from ordinary experience, in particular perception. Epistemological abstraction consists in replacing phenomenal properties with nonphenomenal ones, hence in replacing empirical concepts with transempirical ones. (Recall Section 1.1.) For example, the empirical (subjective) statement "I feel colder now" suggests either of the following transempirical (objective) statements: "The air temperature has dropped" or "There may be something wrong with the subject's health". All mathematical concepts and most of the theoretical concepts in science and technology are epistemologically abstract. Only psychology employs phenomenal concepts, such as that of loudness, alongside transempirical ones, such as sound intensity— and this for the good reason that it wishes to account for subjective experience, not only for objective facts.

The three concepts of abstraction are different but they are logically related. The relation is this: Whatever is either semantically or methodologically abstract is also epistemologically abstract—but the converse is false. All three concepts seem to come in degrees. However, so far only semantical abstraction has been quantitated (Vol. 2, Ch. 6). The quantitation of methodological and epistemological abstraction is an open problem. But at least in a few cases we can order concepts according to epistemological or methodological abstraction. For example, the notions of electromagnetic potential, electromagnetic field intensity, and ponderomotive force, form a ladder of descending epistemological abstraction.

And the concepts of vertebrate, animal, and organism form a ladder of ascending methodological abstraction.

Finally, are there inconceivables, i.e. facts that cannot be conceived of, or concepts that cannot be formed? If there are any we cannot say a priori which they are for, if we did, we would have some conception of them. What we may freely grant is that some facts will forever remain unknown to us, and some conceptual processes are beyond our capacity. But we cannot always specify which ones. Still, we should remember that a number of inconceivables have turned out to be conceivable. Two classical examples are the irrational numbers and change. The former were inconceivable within the Pythagorean doctrine, which held that every number was either an integer or a ratio of integers. As for change, from Parmenides to Bergson a long string of philosophers proclaimed that, although it may be intuited, it cannot be conceived. Yet modern mathematics enlarged the narrow Pythagorean context and moreover proved that there are infinitely many more irrationals than rationals. And modern science places equations of change (motion, field propagation, chemical or nuclear reaction, organic growth, social mobility, etc.) at its very center. So, it is more prudent to talk about *unconceived for the time being* than about inconceivables. And in any case it is more fruitful to try and form new concepts than to speculate about our unknown conceptual limitations.

2. Proposition

2.1. *From Thought to Proposition*

We are now ready to examine complex conceptual operations and, in the first place, thoughts such as thinking that psychology is interesting, that psychology must be allied to sociology, and that all conceptual operations are neurophysiological processes. We have just exemplified three modes of judgment: attribution, association, and generalization. We shall study these basic modes of judgment after the following preliminaries.

Thinking is a brain process the psychological units or modules of which are concepts. Each concept is identical with the specific activity or function of some psychon or plastic neuronal system. (Ch. 2, Section 2.2.) Forming a judgment is thinking of several concepts in quick succession or simultaneously. Therefore judging is identical with the sequential activation of a number of psychons—or perhaps with the simultaneous activation of them. And believing is even more complex, for it is identical to

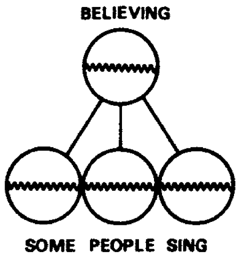


Fig. 5.4. Schematic representation of the sequential activation of the psychons constituting the judgment that some people sing. The evaluating psychon on top does the believing (or disbelieving).

giving assent to such thoughts. Hence presumably believing involves not only conceptual psychons but evaluating ones as well. See Figure 5.4.

Every judgment is a sequence of concepts but the converse is false. Thus the sequence of concepts expressed by the phrase 'She got' is only part of a judgment. We know how to characterize a complete sentence and a well formed proposition, but we still do not know what characterizes a complete judgment. That is, we still have to learn what are the neurophysiological characteristics of a complete judgment as distinct from an arbitrary sequence (or association) of active psychons. This will force us to work most of the time with the conceptual and linguistic manifestations of judgments, i.e. with propositions and sentences respectively.

We may as well start by emphasizing the difference between judgments and propositions. Whereas the former are brain processes, the latter are collections of such and, more particularly, equivalence classes of judgments (Ch. 2, Section 2.2). Consider an external fact such as the song of a bird in a neighborhood. A number of observers may think of this fact, and some of them may agree that it is too early for this time of the year. Most likely the brain processes the various observers go through when thinking this are, though equivalent in some respect, different in other respects. Yet they must be the same in one respect, namely that all of them consist in thinking that the perceived song is too early for this time of the year. And this is precisely the proposition designated by the English sentence 'That song is too early for this time of the year', the Spanish sentence 'Ese canto es muy temprano dada la estación del año', and hundreds of others. Moreover each of these sentences can be uttered in a variety of ways, and likewise each of those judgments can be believed with different intensities. Figure 5.5 summarizes these distinctions, which unfortunately are not always made.

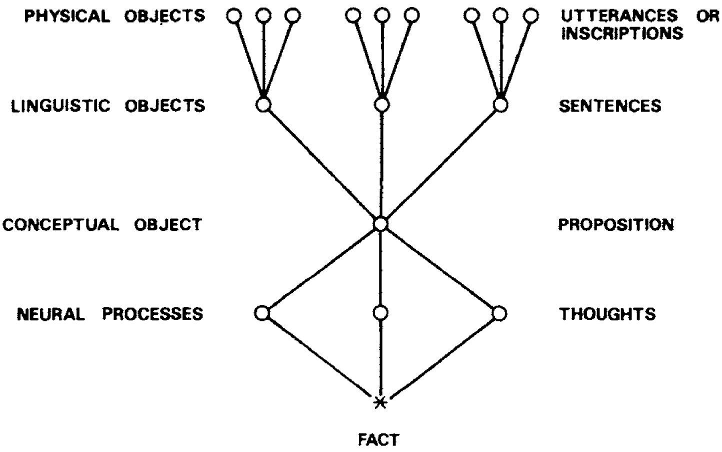


Fig. 5.5. The proposition as an equivalence class of thoughts and the sentence as an equivalence class of utterances or inscriptions.

Some of the categories we have just mentioned are deliberately fused for philosophical reasons. (Fusing is not the same as confusing: the former is conscious equation, the latter is lack of discrimination.) Thus propositions are sometimes called 'facts', perhaps under the influence of idealism. And it is still common to call propositions 'sentences' (and accordingly to speak about the 'sentential calculus') because of the nominalistic prejudice against conceptual objects. Although in most cases such confusions are harmless, in others they are misleading.

(An example of the latter is the problem area of counterfactuals. An entire industry has been built upon the mistaking of counterfactual conditionals for propositions, while actually they are not even sentences. In fact, counterfactuals are ambiguous utterances or inscriptions. Consider the time-worn counterfactual 'If this lump of sugar were placed in water it would dissolve'. This expression may be interpreted in at least two different ways: either as emphasizing that the given lump of sugar is *not* actually immersed in water, or as emphasizing that, when placed in water, it *dissolves*. The reading is a matter of stress or of italics. In the first case we would say 'If this lump of sugar *were* placed in water it would dissolve'. In the second, 'If this lump of sugar were placed in water it would *dissolve*.' Without some information about the context, in particular the speaker's or writer's intention, it is impossible to disambiguate the utterance or the inscription. Because of this ambiguity counterfactual conditionals have no

logical status and do not occur in mathematical or scientific theories. They constitute a pseudoproblem. (See Bunge, 1973b, Ch. 1.) No wonder it is being approached by the bizarre metaphysics of possible worlds, which is not a theory about the only world there is (*Treatise*, Vol. 3, Ch. 4).)

Let us now perform a quick study of the operations of attribution, association, and generalization. *Attribution* is the conceptual operation of assigning a property or a relation to objects of some kind, material or conceptual. Thus when perceiving that b is moving we may attribute to b the property of moving—or, on second thoughts, we may attribute to ourselves the property of moving relative to b. And on perceiving b crashing against c we may attribute to them the relational property of colliding with one another, or even that of attracting each other. (One and the same percept may elicit no judgment, one judgment, two judgments, or even more.)

When attributing phenomenal (observable) properties to perceptible individual things we may not distinguish between perceiving and conceiving, and may even be tempted to espouse phenomenalism. The difference emerges when thinking of properties regardless of their bearers (e.g. "swiftness") or when attributing nonphenomenal properties (e.g. "external"). In either case we obviously go beyond perception and form concepts: in the first case we generalize, in the second we introduce transempirical concepts.

We may conjecture that attributing property P to object b consists in the (simultaneous or successive) activation of the psychons for P and for b. Or we may conjecture that we first think that b is (possesses the property) P, with a single psychon, and then the psychons for b and P are activated (or formed) at the same time or sequentially. Or we may also conjecture that sometimes we undergo processes of the first type (synthesis) and at other times processes of the second type (analysis). We still do not know which, if any, of the three hypotheses is true, and we won't find out as long as they are not spelled out (if possible mathematically) and tested experimentally. However, some crude conjectures have to be hazarded before they can be rendered more precise and put to the test.

Things are easier at the conceptual and linguistic levels, where it is a question of forming propositions or sentences, i.e. of gluing predicates to subjects. Consider a predicate or attribute F representing (conceptualizing) an intrinsic or non-relational property P of objects of some kind K. F can be construed as a function from K to the set of all possible propositions involving F. And attributing P to a member b of K is forming the proposition Fb, i.e. evaluating the function F at b.

Association, in particular pairing, is the general case of attribution. Associations can be natural, as those established between lightning and thunder, or hunger and food; or they can be conventional, like those between side of the road and sense of circulation, or house and street number. And objects of any kind may be thought (truly or falsely) to be associated: things to things, events to events, things to events, concepts to concepts, concepts to properties, etc. Whatever its nature and origin, thought of association may be conjectured to be a neural process similar to that of attribution, i.e. either the pairing of psychons or their sequential activation. The mathematical notion of a relation, or n-ary predicate, elucidates that of association at the conceptual level. Thus consider a relational property P of objects of kinds K and L, such as the relations of connecting, eating, or cooperating. P can be construed as a binary predicate G with domain the cartesian product of K and L, and codomain the set of all propositions containing G. In this case attributing P to the pair of objects b in K and c in L consists in forming the proposition Gbc, i.e. in evaluating G at $\langle b,c \rangle$.

Finally we come to *generalization*. There are two kinds of generalization: (*a*) from like cases, i.e. from individuals to species; (*b*) from one species to another. The former consists in discovering or conjecturing a pattern from individual cases. The second type of generalization is what learning theorists call *transfer of learning* (or of skills)—e.g. having learned one foreign language facilitates the learning of another, or learning carpentry facilitates the learning of work with metals. We shall limit ourselves to generalization of the first type.

How do we generalize? Empiricists hold that every general proposition, in particular every law-statement, is an inductive generalization, i.e. a drawing together of particular propositions. Thus every general hypothesis would come at the end of an inductive process, never at the beginning of a testing procedure. On the other hand rationalists and intuitionists hold that we can start with generalities straight away, on the strength of a single case or even a priori. They also believe that no tests are necessary. Experimental psychologists have found that in fact we and other higher animals proceed sometimes as empiricists and other times as rationalists. Although some generalizations culminate repeated observations, others are formed on the strength of a single case or even prior to observation. That all generalizations, no matter what their origin, must be repeatedly tested before being assigned a truth value, is another matter altogether: here we are dealing with the epistemological problem of the origin or mechanism of generalization, not with the methodological problem of validation.

Consider a classical or Pavlovian conditioning experiment. An animal is presented a light (or a tone) and a short time afterwards is given food (or an electric shock). After a few times the animal has learned to associate the two stimuli, so that on being subjected to the first (unconditioned) stimulus he makes the appropriate response (salivation, withdrawal, or what have you) before the second stimulus is delivered. The light (or sound) need not be always the same, and the food (or painful stimulus) can vary as well: the animal learns quickly to pair a whole class S of stimuli with an entire class R of responses. In mathematical terms, the animal has unwittingly built an entire *function* from S (actually from S cross the set of its internal states) to R out of a few values. (This function is "embodied" in a new configuration of neurons: recall Ch. 2, Section 1.1.) Consequently the animal has acquired a potential or propensity for reacting to further stimuli of the same kind with further responses of the same kind. In short, the animal has generalized—perhaps from a single case. This hypothesis explains instant learning, in particular instant generalization (in contrast to inductive or step by step extension). See Figure 5.6. Incidentally, this hypothesis takes the epistemological aspects of the problem of induction away from philosophy into experimental psychology and neuropsychology. The problem has now become that of identifying the neuronal systems and processes that do the generalizing, either prompted by a handful of cases or prior to observation.

We all know the advantage of generalization: it compresses a large (sometimes infinite) number of epistemic items. (And in the case of the transition of "some" to "all" we also gain in precision.) On the other hand by generalizing we may incur falsity. The whole of magical thinking

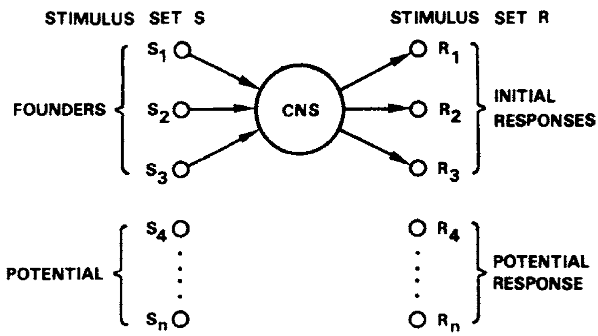


Fig. 5.6. A handful of stimuli are often sufficient to elicit the formation of a new psychon capable of producing new responses of the same kind beyond the initial ones.

derives from hasty (untested) generalization from coincidences. You came just when I was wishing you to come; I was not injured in the accident because I was carrying my amulet; my wrongdoing was not punished because I had paid my tribute to the deity. When such coincidences occur, our nervous system records them; when they do not, it does not. What is wrong about jumping to general "conclusions" is not the jumping itself—a beautiful instance of creativity—but the refusal to subject the generalization to tests before believing it.

2.2. *A Priori and a Posteriori*

What we take to be the basic distinction among propositions was drawn by Hobbes and elaborated on by Leibniz: it is the one between truths of reason and truths of fact. We generalize to half truths (and falsities) and speak in general of *formal* and *factual propositions*. The former are self-sufficient: they do not describe anything real; the latter describe (truly or falsely) some (possible or actual) fact(s). Thus a theorem in pure mathematics is a formal proposition. On the other hand a theorem in some theoretical factual science, like physics or psychology, though proved with mathematical rigor, rests on factual premises because it concerns real things. Besides, its truth or adequacy to fact must be shown by its consistency with the relevant empirical data.

There are several kinds of formal propositions and likewise several types of factual ones. We class them as follows:

- *PROPOSITIONS*
  - *FORMAL*
    - *LOGICAL (T or F under all interpretations)*
    - *MATHEMATICAL (T or F in some model)*
    - *SEMANTICAL (concerning meaning or truth)*
  - *FACTUAL*
    - *HYPOTHESES (axioms, theorems, data) T or F to some extent or degree by virtue of their correspondence with or adequacy to fact*
    - *RULES (prescriptions or procedures). Efficient or inefficient rather than T or F. Refer to human actions (physical or mental) and their referents (material or conceptual)*

The preceding classification is more correct than either of the popular dichotomies analytic/synthetic or a priori/a posteriori. For one thing the concepts of analyticity and aprioriness have not been well defined except in extreme cases. Thus, a tautology is clearly analytic, but what about a theorem in theoretical physics, which has been derived by purely conceptual means? And an empirical datum is clearly a posteriori, but what about a factual hypothesis not built inductively and which happens to anticipate experience correctly? Neither of the two distinctions amounts to a dichotomy, hence neither is a sound principle of classification. When taken together the above classifications produce not only analytic a prioris and synthetic a posterioris but also analytic a posterioris and synthetic a prioris. Our own classification allows us to dispense with all four categories, in particular with the latter.

To be sure, it does make sense to use the expressions 'a priori' and 'a posteriori', but with the understanding that they give rise to subjective and shifting distinctions. Subject A conceives a given hypothesis on the strength of his experience; once conceived, the hypothesis functions as an a priori in A's further exploration of the world. Subject B conjectures the very same hypothesis (e.g. by analogy) and uses it to gain new experience (e.g. by employing it in the design of an experiment). Thus the hypothesis is a posteriori for A and a priori for B at a given moment in each subject's cognitive development; and the next moment what was a priori becomes a posteriori and conversely. In short, the definition "x is a priori iff x is prior to experience and independent of it" leads to contradiction. We must relativize it to the knowing subject and his experience, namely thus:

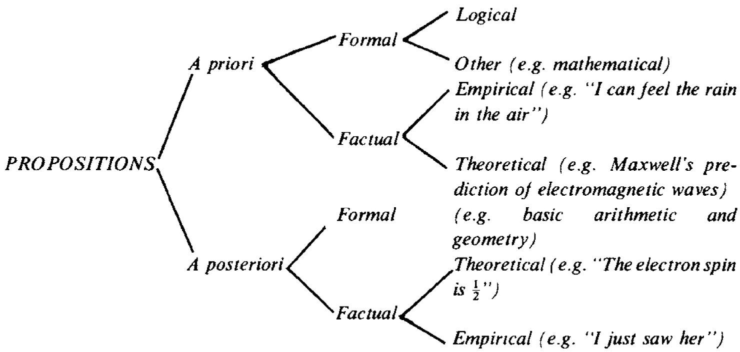
(i) Construct (concept or proposition) x is *a priori* for subject y at time $z =_{df} y$ has formed x by time z, and y has had no experience relevant to x before z (i.e. y knows x without resorting to perception). Otherwise x is *a posteriori* for y at z.

(ii) Construct x is *absolutely a priori* $=_{df}$ x is a priori for any subject at some time. Otherwise x is *absolutely a posteriori.*

These are *epistemological* concepts of aprioriness not to be confused with the *semantic* one of lack of real reference or the *methodological* one of validity regardless of experience. Whereas some mathematical and scientific constructs are epistemologically a priori, only the formal ones are methodologically a priori, i.e. they stand or fall regardless of any empirical operations. Absolutely a priori constructs, such as those of logical consistency, infinity, and topological space, are not to be confused with

innate ideas of the kind postulated by Socrates, Freud, or Chomsky. In particular the constructs of logic and mathematics, though seldom originating in experience, cannot be formed without considerable prior experience of some kind. Only our nervous system, possessing certain dispositions built into its very organization, is innate. Thus it is possible that the tendencies to locate external events in space and time, and to search for causal connections, are innate, though perhaps not inborn. However, such dispositions do not constitute knowledge: they are only necessary conditions for learning.

The a priori/a posteriori dichotomy can be elaborated on. In particular, we must admit *factual a prioris*, such as the novel predictions made with the help of some scientific theories. (Think of the "missing links", now found and called *hominids*, or of the forecasts of World War II.) The resulting classification is this:



3. Exactification

3.1. *From Vagueness to Exactness*

Many, perhaps most of the concepts designated by ordinary language expressions are vague, i.e. imprecise or inexact: think of "small" and "most". Even in mathematics and science newly minted concepts are likely to be somewhat vague, and none becomes completely exact unless incorporated into a theory. Whereas in some cases vagueness is harmless and even expedites thinking—or at least speaking—in others it is a serious stumbling block to thinking and communicating.

Vagueness precludes the use of logic, which deals only with exact

predicates. (In particular vague predicates violate the excluded middle principle, which rejects borderline cases by stating that nothing has and fails to have any given attribute.) Take for example the concept "most", the negation of which is "none or some or all". Now, 'none or some" is tautological, for it amounts to "not some or some". Hence the negation of a most-statement amounts to the disjunction of a tautology and a universal statement. So, no matter what the truth value of the latter may be, the disjunction is logically true—which renders some-statements immune to tests. This shows that the 'most' of ordinary language violates logic. But the concept can be exactified, e.g. by construing "Most A's are B's" as "The relative frequency of A's that are also B's exceeds $\frac{1}{2}$". This construal avoids the above paradox.

Vagueness can often be reduced or even eliminated by some conceptual means or other, such as definition, quantitation, or theorization. Thus the expression 'B depends on A' can be rendered exact by construing A and B as sets, or members of sets, and dependence as a mathematical function from A to B. (For example, the ontological thesis that the past determines the present may be exactified as "For every thing, and every property of it relative to any given reference frame, there is a function representing that property, with respect to that frame, and such that the current value of the function is a function or a functional of its past values".) The vague concepts of fatness and baldness may be replaced with a certain standard weight value and a certain standard number of hairs respectively. The indefinite "some" may be eliminated either by quantitating or by dividing the original set into homogeneous subsets every member of which possesses the property designated by the predicate in question. And the ever present "etc." may be replaced with either an exhaustive list—which is not always possible—or a mention of the proximate genus. Thus instead of saying 'spinach, broccoli, string beans, etc.', we can say 'green vegetables', thereby gaining in both precision and breadth. Being a conceptual defect, vagueness can be reduced only by conceptual means: purely linguistic tricks will not help.

Vagueness is imprecision in meaning. Now, meaning can be construed as composed of sense and reference or, in self-explanatory symbols, $\mathscr{M}(c) = \langle \mathscr{S}(c), \mathscr{R}(c) \rangle$ (Vol. 2, Ch. 7). The sense of a construct in a given context equals the set of its logical relatives, i.e. implicants and implicates. And its reference class is of course the set of all the objects referred to by the construct. (Caution: Reference must not be mistaken for extension. A predicate referring to mythical objects refers and yet has an empty

extension.) Since the meaning of a concept is composed of its sense and its reference, its vagueness can be of either. We shall therefore speak of intensional and referential vagueness. Moreover an imprecise sense may result in a somewhat indefinite reference class: think of the concepts of thing, organism, thought, and democracy.

The concept of *intensional vagueness* can be analyzed in terms of the full but unknown sense $\mathscr{S}(c)$ of a construct c, and the subset $\mathscr{S}_k(c)$ of known members of its sense, namely thus: $\mathscr{V}_i(c) = \mathscr{S}(c) - \mathscr{S}_k(c)$. But, since only $\mathscr{S}_k(c)$ is well determined, the concept of intensional vagueness is itself vague and therefore of little use. See Figure 5.7. On the other hand the concept of referential vagueness is quite clear-cut. A reliable test for the referential definiteness of a predicate is its performance in dichotomy: if a predicate allows us to unambiguously and exhaustively perform a partition of any set into the subset of all the members with the given property and the complementary subset of all those elements that lack it, then the reference class of the predicate is entirely definite. If, on the other hand, there are borderline cases that might as well be grouped with one category or its complement, then the predicate is referentially vague.

(A precise measure of such vagueness is provided by the fraction of borderline cases. In obvious symbols, $\mathscr{V}_r(c) = n/N$, a formula that works only for finite universes of discourse. It follows that $\mathscr{V}_r(\neg c) = \mathscr{V}_r(c)$, i.e. referential vagueness is shared to the same extent by a concept and its dual. And the *degree of definiteness* of c may be defined as the complement of that number to unity. Warning: These are just the degrees of referential vagueness or definiteness *known* at a given moment to a given knowing subject. Different workers may and, in fact, usually do, come up with different counts of borderline cases—a common occurrence in ordinary life as well as in biological and sociological systematics. So, although the concept of referential vagueness is quite precise, its application varies from

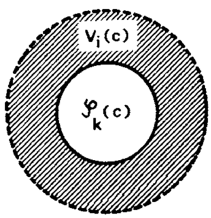


Fig. 5.7. Intensional vagueness: the ill-limited shaded zone.

subject to subject. Finally, the *total vagueness* or *indefiniteness* of a concept c may be defined as $\mathscr{V}(c) = \langle \mathscr{V}_i(c), \mathscr{V}_r(c) \rangle$.)

Intensional vagueness is a necessary but insufficient condition for referential vagueness. I.e., if a concept is referentially vague then it is also intensionally vague but the converse need not be true. In fact intensional vagueness can often be palliated by introducing and using practical *criteria* (decision rules) enabling one to decide whether a given object belongs to the reference class of the concept in question. Thus long before any of the essential properties of gold were known, chemists used *aqua reqia* for identifying golden things. And, although nobody knows what scientific talent is, anybody can spot it by counting numbers of tough solved problems, or even publications.

However, in scientific contexts we cannot be satisfied with referential precision: we also need intensional definiteness. This is best achieved by turning the given context into a member of a mathematical theory. However, mathematization has its limits: (*a*) it is only necessary, not sufficient, as shown by the case of fuzzy set theory, and (*b*) it concerns only the form or structure of a concept, not its factual reference. In any event, regardless of the means whereby a construct is sharpened, we may stipulate that a construct is *exact* if and only if it is not vague—i.e. if both its intensional and its referential vagueness are nil.

To *exactify* a concept, then, is to replace it with another having a precise sense and a precise reference, each of which has a sizable overlap with the sense and the reference, respectively, of the original imprecise concept. A perfect coincidence of sense and reference cannot be expected: every time a concept is exactified something is gained and something else—usually simplicity and intuitiveness—is lost. Think of the contemporary concepts of number, matter, and life, as compared with their predecessors.

Needless to say, exactification is not identical with quantitation: there are numerous qualitative mathematical disciplines, such as logic, abstract algebra, and topology. Nor is exactification restricted to the natural sciences: there is a vigorous exactification movement in the social sciences and the humanities, not excluding philosophy. (See, e.g. Bunge (Ed.), 1973d.)

Concept refinement, or exactification, is one aspect of the progress of knowledge. In fact the latter does not consist in piling up new data and new conjectures, but in a selective process that pivots around the invention and improvement of concepts, propositions, theories, and methods. In this process concept refinement plays an indispensable part: although it does

not replace the original invention of an initially imprecise idea, it helps to develop it. This must be insisted on because the role of concept refinement has been exaggerated by some and underrated by others. The overrating has been mainly the work of theorists, which is understandable, for they are professional concept breeders. And the underrating is not surprising either in view of the empiricist tenet that genuine research is concerned only with fact finding. This one-sided view was justified at the beginning of the modern era, as a reaction against the hollow verbosity of decadent scholasticism. Recall that Bacon advertised his own philosophy as "a philosophy of works, not of words", and that the Royal Society of London took from him its motto *Nullius in verba*—"There is nothing in words".

The anti-intellectualist tendency became obsolete no sooner it was born, for modern science and the technology based on it are as conceptual as they are empirical. In fact many problems in science and modern technology are problems of concept formation or refinement, which occur in the building, application, and evaluation of theories. Think of the two centuries of scientific debates over the term 'force', which turned out to designate now the concept of force proper, now that of energy, now that of power. Or, to turn to present day problems, think of the notions of mind, creativity, and intelligence, which still roam wildly in psychology. Or of the concepts of economic value, money, and social class, paragons of inexactness in social science. The vagueness of these notions hampers theoretical development and empirical research: it only encourages confusion and barren argumentation. (See an unabashed praise of imprecision in Feyerabend 1981, p. ix.)

What can be done about inexact concepts? The tradition in mathematics and natural science has been to replace them with exact concepts. In recent times a new strategy has been advocated, namely that of accepting vague predicates and treat their ill defined extensions as fuzzy sets. (An exact predicate has an exact extension, i.e. one the membership of which is definite. An inexact predicate has a fuzzy extension, i.e. one the membership of which is not well defined; hence, given an individual, one may not be certain whether or not it belongs to the set. The mathematical theory of fuzzy sets has been invented to cope with such collections.) The decision to use fuzzy sets or abstain from using them in representing reality is a methodological one. The usual arguments are that reality itself is fuzzy, and that fuzzy concepts are respectable after all since we can speak about them in an exact way with the help of the theory of fuzzy sets. Let us deal briefly with these two arguments.

There is no evidence that reality itself is fuzzy or indeterminate. What is

true is that certain exact concepts do not apply to some things: for example, the concepts of position, trajectory, and shape do not refer correctly to elementary particles. In such cases one introduces different but still exact concepts, such as that of position distribution instead of sharp position. Moreover all the successful factual theories have formalisms presupposing the theory of ordinary sets. In short, it is not that the world is fuzzy, but that our knowledge of it is imperfect (but perfectible). As for the claim that fuzzy sets have become respectable for being the subject of a mathematical theory, it is like saying that crime became honorable since the inception of criminology. Using fuzzy sets in modeling reality is a confession of defeat. The point is not to be able to speak exactly about inexactness but to reduce the latter and, if possible, to eliminate it altogether. Dirt must be recycled or eliminated, not be shoved under the carpet, let alone exhibited proudly. "Everything that can be thought at all can be thought clearly. Everything that can be said can be said clearly" (Wittgenstein, 1922, 4.116).

3.2. *Quality and Quantity*

Primitive and archaic thought are characteristically qualitative and analogical (Lloyd, 1966; Goody, 1977). By employing exclusively qualitative or dichotomous predicates, primitives and ancients were necessarily led to dichotomous or polar classings: cold and hot, heavy and light, clear and dark, near and far, wild and domesticated, good and bad. This mode of conception misses nuances and is apt to lead to a Manichean or a dialectical world view. Scientific and technological thinking, on the other hand, presumes that most properties come in degrees—of heaviness and speed, heat and nearness, excitation and value, and so on. Only in some backward branches of the social sciences and the humanities most thinking remains qualitative, hence polar or black and white.

True, a few properties are objectively qualitative, hence correctly representable by dichotomous predicates. Existence is an, perhaps *the* example: a thing either exists or it does not—there are no degrees of existence. (For the construal of existence as a predicate see Vol. 3, Ch. 3.) But most, perhaps all other properties, come in degrees that can be ordered or even assigned numbers. For example, there are different degrees of stability, including metastability and ultrastability. And, of course, there are infinitely many degrees of size, speed, probability, and what have you.

We cannot know a priori whether a given property is inherently qualitative or will turn out to be quantitative. We can only try to form a

quantitative conceptualization of the given property and measure the latter to see whether we get more values than one (presence) and zero (absence). In any case modern science and technology since Galileo have adopted the methodological postulate that *every property, except for existence, can be quantitated.* Faith in this postulate has sustained all the investigators who have succeeded in quantitating properties previously believed to be imponderable. The postulate is confirmable but irrefutable for, should a given property resist our efforts to quantitate it, we may blame ourselves for our failure.

Having praised quantitation we must warn against the temptation to grade even complex properties by single numbers. Such oversimplification results in adopting IQ as an overall measure of intelligence, and increase in GNP as an overall index of progress. When such simple-minded quantitations are shown to be inadequate, irrationalists are confirmed in their mistrust of numbers. Better not give them this pleasure and recognize that in many cases a scalar will not do, but that a vector, a tensor, or a matrix may be needed.

Evidently quantitative knowledge, when available, implies qualitative knowledge but not conversely. Thus "distance" subsumes "separateness", "rate of change" subsumes "changing", and "probability" subsumes "possibility". If a quantitative pattern (law) has been found, its qualitative features can be extracted. Thus whether the evolution of a system is stable or unstable, can be found out theoretically only after its equation of evolution has been set up and tested—which equation involves numerical functions. (This is how Poincaré proved the stability of the solar system.) Therefore it is not true that the study of qualities always precedes that of quantities. The deepest qualitative knowledge is a byproduct of some quantitative knowledge. All of the social sciences could profit by this maxim; some of them, in particular economics and socioeconomic history, are already applying it.

The fundamentals of quantitation are these. Let S be a set of degrees or intensities of some property—objective degrees such as temperatures or subjective ones such as hotness. Postulate that S is simply ordered by a relation $\gtrsim$. (Hence if x and y are in S, then $x \gtrsim y$ or $y \gtrsim x$, and both statements hold only in case $x \sim y$.) In most cases this comparative concept is insufficient. We *quantitate* the given concept by mapping S onto a set of numbers in such a way that (*a*) each degree (member of S) be assigned a single number (one to one correspondence), and (*b*) the order in S be preserved, i.e. the numerical image of S reproduces the objective order in S.

(In other words the quantitation of S consists in introducing a mapping $M:S \to T$, where T is included in the real line, such that, for any x and y in S, $x \gtrsim y$ if and only if $M(x) \geq M(y)$. Thus "farther away than" will be quantitated as "at a greater distance than", and "hotter than" as "having a greater heat content than". In general such mapping is not unique: i.e. in principle there can be different quantitations of one and the same property. Think of the different temperature scales and the different modes of reckoning volumes of production. The choice among the various possibilities is dictated in some cases (e.g. length) by convenience, in others (e.g. temperature) by theory. If S is or is assumed to be nondenumerable, so must be its image T even if we do not possess the technical means for measuring the difference between two close members of T. If the map M is one to one, order preserving, continuous, and has a continuous inverse, it is called a *homeomorphism*.)

(The above procedure presupposes that we are given the entire set S of degrees of some property. This is sometimes the case. As a rule we build a quantitative concept regardless of empirical findings, and design observations or measurements in order to find some of the values of the function concerned. More precisely, we introduce a function $M: D \to T$, where D is in general a cartesian product of certain sets. At least one of these is a set of concrete things, such as bodies or societies, and at most one other factor is a set of units, such as the multiples and submultiples of the meter. For example, if we wish to reckon or measure distances between things in the context of relativistic physics, we introduce the distance function $d: P \times P \times F \times U_d \to \mathbb{R}^+$ obeying the usual distance axioms, where P is a set of physical objects, F that of reference frames, U_d the set of distance units, and $\mathbb{R}^+$ the set of positive real numbers. The mathematical theory of metric spaces leaves P uninterpreted and fails to include F and U_d. Any map $M:D \to T$ with the above characteristics is called a *quantitative predicate* or *magnitude*. See Bunge (1971, 1973a).)

(Qualitative predicates are particular cases of quantitative ones. Indeed the former are functions of the form $Q: D \to \{0, 1\}$ such that $Q(x) = 1$ iff x possesses the given property, and $Q(x) = 0$ otherwise. Once we have a set of magnitudes, e.g. those occurring in a mathematical model of a real system, we may focus our attention on certain global or qualitative features of the system. To this end we may need to *discretize* those magnitudes, i.e. to replace them with functions ranging over the integers. A common discretization is that of time, particularly when computers are employed. In this case differential equations are replaced with finite difference equations.

The discretization of a continuous system model is not always a trivial exercise. Sometimes it is suitable to discretize the state variables themselves, as when one stipulates that any values below threshold count as 0, and any above threshold count as 1. At other times one is interested in growth and decline rather than in absolute values. In such cases the rates are discretized, namely thus: the state of the variable is assigned the number 1 if its rate is positive, and 0 if negative. In this case the momentary state of the system is represented by an n-tuple of zeros and ones exhibiting the state of each variable: Glass (1975). In short, once again the qualitative features of the system are inferred from a quantitative model of it.)

Every real property is representable by at least one mathematical function, but the converse is false. Genuine quantitation obeys a methodological condition as well as a mathematical one. It must be accompanied by an indication that there are (direct or indirect, actual or potential) ways of effectively assigning (measuring) some values of the function in question. Otherwise quantitation will be judged phoney. Three examples should suffice. One is Herbart's phoney mathematical psychology. Another is the widespread belief, among contemporary philosophers, that propositions can be assigned probabilities, even though they are very discreet with regard to the method whereby such assignment is to be made. A third example of phoney quantitation is this. One can easily speculate that dreams have densities, and can imagine some postulates defining the dream density function—e.g. if dream a precedes dream b, then the density $d(a)$ is greater than the density $d(b)$. But this number game has no substance because we do not have the faintest idea how to determine effectively $d(a)$ and $d(b)$.

So much for quantitation, which we shall meet again when studying measurement (Ch. 11, Section 3.1). The two operations should not be confused, although they often are in the methodological literature (e.g. in Suppes and Zinnes, 1963). Quantitation is a purely conceptual operation, whereas measurement is empirical as well as conceptual. We can have quantitation without measurement but not conversely. For example, the concept of power exerted by organized crime in a given society may be quantitated, e.g. as the number of people whose economic or political activities are controlled by it, or by the annual volume of its business. However, since organized crime is supposed to be illegal, no reliable statistics can be compiled to measure its power. Still, this power is real, it can be roughly estimated, and one day it may be discovered. On the other hand it is impossible to measure in any objective way the probability of

propositions, in particular hypotheses, so assigning them probabilities is a prime example of bogus quantitation. Although quantitation is different from measurement and precedes the latter, it is pointless unless measurement is possible at least in principle.

4. Representation

4.1. *Conceptual Map*

Percepts tend to cluster into systems that map the body or the environment: those systems are the perceptual maps, which allow the animal to move around, make a living, and defend itself (Ch. 4, Section 3.1). Concepts also tend to organize into systems: some of them map the body, others the environment, and finally some (namely the theories in formal science) map nothing at all. All animals endowed with a central nervous system form perceptual maps, but only the higher vertebrates seem capable of forming conceptual maps, i.e. representations that overreach perception. Unlike the perceptual maps, the conceptual ones allow the animal to understand, forecast, and plan to some extent.

All conceptual maps are patterns of neural activity. However, some are externalized and thus become perceptible to other subjects: this is the case with diagrams, geographical maps, and systems of sentences or utterances. Whether or not externalized and thus visualizable or audible, conceptual maps summarize and systematize conceptual knowledge (or prejudice). And, whether or not they are accurate (true) representations of real items, they help us orient ourselves (successfully or not), as well as to act in the real world, in at least as effective (or misleading) a manner as perceptual maps.

Consider diagrams for a moment, and in particular the diagrammatic representations of a concrete system with known of conjectured composition, environment, and structure. About the simplest such representation is a diagram exhibiting the components (at a given level), the (relevant) environmental items, and the (main) relations among all such items: see Figure 5.8. Note that, although the diagram is perceptible, it is a conceptual or symbolic map not a perceptual one, if only because it may represent imperceptible things, such as an atomic nucleus composed of two nucleons held together by imperceptible links, such as fields. Note also that such a diagram is an extremely poor representation of any system: it summarizes only the minimal body of knowledge necessary to speak about a system. A stack of diagrams of the same type may provide a more detailed

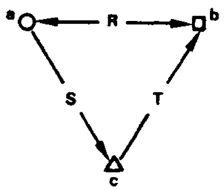


Fig. 5.8. Schematic representation of a system with two components, a and b, held together by the link R, and interacting with environmental item c via couplings S and T.

representation. But only a theoretical model of it—e.g. a system of differential equations—can give a detailed and deep representation. (We shall deal with theories in Ch. 9.)

Here are a few principles concerning the diagrammatic representation of material things:

(i) A diagram of a system should represent its composition, environment, and structure, either directly or in terms of properties. (A mixed diagram, containing arrows from things to properties or conversely, makes no sense because one and the same symbol, namely the arrow, is now made to denote a link or coupling, now the subject-predicate relation. However, such diagrams occur frequently in the literature.)

(ii) One and the same concrete system may be represented in alternative ways, and any such alternatives may or may not be equivalent. (They are so only if they can be obtained from one another via a topological transformation, i.e. one preserving the relations of vicinity.)

(iii) If two diagrams representing a given system are inequivalent, then they complement each other or one of them is more accurate than the other.

(iv) Diagrams are no substitute for theories. In particular, although they may suggest explanations, they cannot produce forecasts.

Like perceptual maps, conceptual maps can be analyzed as partial functions from stimulus events to neural events. (Cf. Ch. 4, Section 3.1.) But, whereas perceptual maps are limited to appearances, conceptual maps may represent reality, i.e. they may be subject-invariant. Think of a coin, its possible appearances to a subject, and the conceptual models of it. A single coin may appear to us in any one of infinitely many fashions—as many as its relative positions to ourselves, dynamic states, illumination conditions, etc. On the other hand only a finite number of conceptual models (e.g. geometrical, dynamical, and probabilistic) of a real coin can be built, none of which involves a subject.

Another difference between the two kinds of map is that, whereas perceptual maps resemble the represented objects, conceptual maps need not hold such a relation of analogy with their objects or referents. Think of a family tree, or of a system of econometric equations: unlike blueprints and geographical representations, such symbolic maps do not resemble their referents. Therefore the words 'picture' and 'reflection' do not describe them adequately. This suffices to write off naive realism (the copy or picture or mirror view of knowledge) as inconsistent with science and technology.

Contrary to naive realism, Helmholtz held that all our representations are symbolic and that even perceiving involves conjecturing. He wrote that representations are not copies or images but "only *signs* [symptoms] of external objects, and in no sense images of any degree of resemblance. [...] For a *sign* it is sufficient that it become apparent as often as the occurrence to be depicted makes its appearance, the conformity between them being restricted to their presenting themselves simultaneously; and the correspondence existing between our sensations and the objects producing them is precisely of this kind. They are signs which we have *learned to decipher*, and a language given us with our organisation by which external objects discourse to us—a language, however, like our mother tongue, that we can learn only by practice and experience" (1873, pp. 391–2).

Conceptual maps, and particularly factual theories, are indeed symbolic in some respect or other and so need not resemble their referents: Helmholtz's view is true of them. On the other hand it does not hold for perceptual maps: if the latter did not preserve at least roughly (some of) the relations among the stimulus events, we would be unable to move around and act in our environment. So, we accept Helmholtz's view for conceptual maps but not for perceptual ones. Incidentally, that view, though at variance with naive realism, is consistent with critical realism and even with materialism, for it presupposes the autonomous existence of the external world: Helmholtz insists that our ideas are signs of the latter. (This point was misunderstood by Lenin (1908, pp. 237ff.), who thought that if our ideas are only signs then the existence of external objects becomes doubtful.)

Naive realism is allied to mechanism, according to which the world is a mechanical system, so that every real thing can be pictured literally, i.e. described "*par figures et mouvements*" (Descartes). Since the inception of field physics we have learned that this view is overly optimistic; quantum mechanics has weakened further the mechanistic thesis, and the social

sciences have never had any use for it. We know that only some physical objects, in particular medium sized mechanical systems, can be represented iconically, and even so only in a few respects. Electrons and photons, biopopulations and societies, cannot be diagrammed in detail: only symbolic systems can represent them. This does not mean that they are less real than stones, but that they have properties that cannot be grasped by perception. But, of course, the theories representing such things can be externalized by means of written symbols, and thus become vicariously perceptible.

Once a conceptual map has been drawn, written down, or taped, it becomes a cultural artifact, i.e. one that can be examined, modified or used by others. By becoming public, the map seems to acquire an existence of its own. But this is an illusion, for the drawing, printed page, or tape, would not perform any functions in the absence of brains. In other words, we have to do with three objects, and therefore three relations: the represented thing, its conceptual map in some brain, and the public manifestation of the latter. The relation between the latter and the thing it maps is a sort of product or composition of the relations artifact–brain, and brain–represented thing. See Figure 5.9.

4.2. *Discovery or Invention?*

If conceptual maps are not mere copies but instead symbolic constructions, then they must be creations like poems or musical scores and unlike photographs or tapes. Hence when exploring our environment we do not always make discoveries, as when finding a coin in a pocket, but sometimes we invent beforehand or at the same time that we discover, i.e. we discover a thing and invent a conceptual map of it. Thus Columbus was able to discover new lands for the Europeans only because he had previously

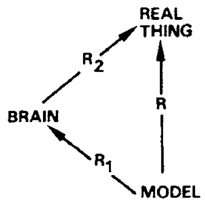


Fig. 5.9. The modeling relation R equals the composition of the model-brain and the brain-represented object relations.

formed a (rather inaccurate) conceptual map of the planet. When Harvey announced that he had discovered the circulation of the blood, he built a conceptual map of the human cardiovascular system complete with capillaries, which he had been unable to see. (So, the inimitable 17th century gossip John Aubrey (1949) was not totally wrong when he described Harvey as "Inventor of the Circulation of the Bloud"—as a result of which "he fell mightily in his Practize, and 'twas beleeved by the vulgar that he was crack-brained.") And when J. J. Thompson discovered the electron in 1898, he produced a conceptual map of it as a tiny electrically charged corpuscle. Others might have thought of cathode rays as jets of an electrically charged fluid: they would have missed discovering the electron.

Every important discovery is accompanied by the invention of a conceptual map of the object discovered. Moreover such a map is what makes the discovery important. Thus if an ignorant person finds a primitive tool and does not recognize it as such, he makes one more discovery among the many he makes every day. If the same object is found by a competent archaeologist, he may fit it into a conceptual map of a primitive society, and thus turn it into an important discovery. Likewise when an inventor imagines new uses of known things, assembles old units into new systems, or designs totally new systems, he forms conceptual maps. And some of his inventions may help make new discoveries. What holds for the invention of artifacts holds also for mathematical invention. Even the discovery of a previously unknown consequence of a set of premises, as in the proof of a theorem, is a creative process, for it leads to something new. A fortiori, the invention of new concepts, new theories, or new theorems is a creative process, and moreover an original one.

People are creative all the time: they keep on discovering or inventing things or ideas, mostly ordinary, sometimes extraordinarily clever or silly. But not all creations are original: many are rediscoveries or reinventions. A creation is said to be *original* when nobody but its creator knew it before. Obviously, original creations are not equally valuable: whereas some are good, others are bad and most are indifferent. Hence originality is not good in itself: only good originality is. Nevertheless some societies encourage originality in itself whereas others discourage originality and thus block progress—which is understandable, because novelty may threaten any rigid social order.

The notion of creativity is suspect in many quarters, in particular in empiricist philosophy and its psychological counterpart, behaviorist psychology. According to these doctrines all the organism can do is to

admit environmental inputs, in particular pick up the information inherent in the environment. This environmentalist thesis would never have occurred to a historian: he knows that theories and machines are invented, not gathered like wild fruit or fish. Nor does the neuropsychologist have any difficulty in admitting creativity: to him it is the ability to form new neuronal configurations (Bunge, 1980a). This ability is partly innate, partly fashioned (or deformed) by education. A dogmatic education will systematically punish originality and encourage blind repetition, whereas a progressive education is supposed to free creativity—though it sometimes fails to promote it for fear of falling into dogmatism.

In general, all the higher vertebrates are creative, and a few individuals are original in some way or other. There is nothing mysterious about originality although there is much to learn about it. Every one of us has a peculiar brain composition and organization, and every one of us is subjected to environmental circumstances that differ somewhat from those of other individuals. (In particular, no two individuals have exactly the same protein composition and consequently can engage in exactly the same neurochemical reactions. And no two humans are treated exactly alike by their fellow humans or are presented with exactly the same opportunities and challenges.) In conclusion, there is a grain of truth in empiricism: we cannot escape our environment although we can modify it somewhat. And there is a grain of truth in rationalism as well: we create our own ideas, in particular conceptual maps—though not out of the blue. So, the whole truth contains a grain of empiricism and another of rationalism. However, it is not a mere merger or synthesis of the two, for each of them ignores the brain and that very particular environment which is society. We shall return to this subject in Ch. 15.

We value creativity but do not know the secret of it. Nobel laureate Linus Pauling was once asked how he had managed to propose so many good original ideas in so many different fields. He replied: "It's simple. You just have plenty of ideas, and cast off the bad ones". Yes, but how do you come by plenty of ideas, good or bad? And how do you manage to spot the bad ones? The first question is yet to be answered by the psychology of thinking; the second will concern us in Ch. 12.

5. Concluding remarks

Perception gives us only perceptual knowledge, which is egocentric and limited to appearances. Only conceptual knowledge can be objective

and deep: only conceptual maps give us a glimpse of things in themselves. (Cf. Schlick, 1925.) Still, conceptual knowledge is not independent of perception, which sets conception in motion and checks its products. This view, which can be found *in nuce* in Aristotle, is part of every critical realist epistemology.

Critical realism takes some elements from empiricism and others from rationalism but goes beyond both of them. Empiricism has always extolled the virtues of perception and rejected all concepts that did not originate in it. On the other hand rationalism has always praised the superiority of conception to the point of proclaiming the self-sufficiency of reason. Critical realism is a sort of synthesis of empiricism and rationalism, but it is not the only possible synthesis. Recall that Kant had managed to put together the negative aspects of empiricism and rationalism, by holding that we can have no experience without certain a priori intuitions, that things conform to human thought rather than the other way round, and yet we can know only phenomena, not things in themselves—all of which is out of tune with the realistic epistemology inherent in modern science and technology. (More in Ch. 15, Section 3.)

Although we pride ourselves on being the only animals capable of certain conceptual feats, such as building mathematical theories of the world, we still know very little about the psychology and neurophysiology of conceptual operations. This backwardness seems to have two main causes. One is our familiarity with such processes. (Familiarity may not breed contempt, but it certainly fails to spur inquiry.) The other is the belief that those operations could not possibly be studied scientifically—a belief nurtured by behaviorism. However, at last the scientific study of conception is under way since about 1960. Besides, we have learned to analyze many conceptual processes and have set up a number of rules for the correct performance of some of them, such as classification and deduction. Paradoxically, methodology is somewhat more advanced than descriptive and explanatory epistemology: we know how to perform certain operations correctly while having only very rough ideas of how we manage to do it.

Conceiving is anything but copying: it is a creative process going far beyond perception—which in the higher vertebrates is creative too, partly because it gets mixed with conception. We know very little about the mechanisms of conceptual creation, and this partly because of the empiricist tenet that every conception is a mere combination or elaboration of percepts. The classical examples are those of the centaur and the

mermaid. The myriads of counterexamples supplied by logic, mathematics, science, technology, and the humanities, are merely ignored. Yet the invention of syllogisms, equations, grammars, the idea of an atom, and the wheel, owe nothing to perception. All we know is that conception is a brain process: we ignore its mechanism, and a fortiori the secret of originality.

CHAPTER 6

INFERRING

Conceptual knowledge can be augmented in either of two ways: by forming new propositions out of nonpropositional material (e.g. percepts), or by inferring new propositions from a set of known ones. We do the former, for instance, when producing empirical data and proposing conjectures to explain them. And we infer when, for example, drawing conclusions from a set of premises and leaping to generalizations on the strength of particulars. Both processes are equally important in all fields of inquiry. However, the importance of inference is often underrated by those workers who, misled by empiricism, believe that all they have to do is collect solid facts and abstain from committing what they regard as the capital sin of the intellect, namely jumping to conclusions. But, of course, those who do not jump do not overcome any hurdles, and so do not participate in any original cognitive venture.

As in the case of propositions, it is possible to adopt either a psychological or a philosophical (and in particular logical) viewpoint concerning inference. That is, we may regard it either as a brain process without paying attention to its validity, or as a timeless relation between propositions and focusing on its validity with total disregard for its neurophysiology. In line with our program of fusing the scientific and the philosophical approaches to inquiry, we shall adopt both viewpoints. However, we shall spend far less time on the former than on the latter because we know very little about the physiology and psychology of inference.

Every inference is either deductive or nondeductive. The former is demonstrative, the latter nondemonstrative and at most suggestive. Deduction is of course the central theme of logic, which has become a separate branch of mathematics. We shall have little to say about it. Nor shall we spend much time on nondemonstrative inference, in particular induction and analogy. Not that induction and analogy are unimportant, as the rationalists claim: they are pervasive, though perhaps not more so than intuition. Moreover they can and must be studied, though not with the hope of coming up with exact rules or methods, such as the rules contained in works on inductive logic and the much-vaunted inductive method.

Likewise love and hate deserve being studied, though not with the aim of producing scientific love or scientific hate.

The above does not entail that all nondemonstrative inferences are worth the same, namely little, let alone nothing. Nondeductive inferences, like guesses, can be wild or educated, gratuitous or compatible with data or theory. But whether or not they happen to be lucky is for tests, not for an a priori discipline, to decide.

1. From Spontaneity to Formality

1.1. *Natural Reasoning*

The psychobiological investigation of reasoning patterns is a recent development which is yielding few but important results on the thinking habits of children, primitives, and adults living in modern societies. (See Nisbett and Ross, 1980.) We shall make a quick review of some typical findings.

In all inference we must assume some premises, even if we do not know whether they are true, if only for the sake of argument, in particular to find out whether such assumptions lead to acceptable consequences. This detached attitude is not universal. The natural thing to do is either to accept or to reject a premise rather than to examine it with a purely cognitive concern. Entertaining or examining assumptions without committing oneself a priori to them is not done in early infancy and may be a rather recent historical acquisition. At least, the historical record does not seem to go farther back than a few centuries before our era.

In particular, children do not seem to be able to form conditionals, i.e. propositions of the form "If A then B", until about age 3, and this only when prompted by their teachers. Moreover when they do form a conditional they presuppose its antecedent A to be true: it is only much later that they can envisage conditionals with false antecedents or antecedents with unknown truth values. Likewise nonliterate peoples seem incapable of forming such conditionals. (Cf. Luria, 1976.) In short, hypothetical reasoning comes rather late or not at all: it seems to be peculiar to an advanced stage of culture.

Not surprisingly, argument too comes rather late. Piaget (1955) found that children go through two stages with regard to argument: quarrelling and argumentation proper. The former consists in a clash of views, and it is the mental equivalent of fighting. Genuine argument, which exhibits

reasons, motives, or causes, does not appear until age 7. (But, presumably, the majority of such rational arguments are formally invalid.) Another result of Piaget's studies (1955) is that, before the ages of 7 or 8, children may reject self-contradictions but they do not often notice contradictions between opinions voiced successively: their body of beliefs does not form a system. Between 7 and 8 they start noticing and rejecting contradictions, thus building gradually belief systems or world views. (Whether this process is universal or peculiar to advanced cultures, is not known.)

Once a person has attained this rational stage she tries to eliminate contradictions within her body of beliefs by discarding some of them. However, if such contradictions persist—if the person cannot give up one of the theses in conflict—the individual tries to "rationalize" them. If she fails, she feels a psychological discomfort. And, as long as she feels such discomfort, she acts so as to reduce "cognitive dissonance", much in the same way that she tries to reduce other imbalances, such as hunger. One way of reducing dissonance is to avoid situations where the individual is forced to choose between the horns of her dilemmas. Social intercourse is at the same time a source of cognitive dissonance and a vehicle for eliminating it by introducing new knowledge. The greater the dissonance, the stronger the social pressure to reduce it. (Festinger, 1957).

Although the theory of cognitive dissonance does not seem to have been established experimentally (Chapanis and Chapanis, 1964), casual observation confirms that ordinarily we abhor contradictions but, when we cannot get rid of them, we atone for them in some way or other. For example, the contradictions inherent in every religion lead some people to claim that they are mysteries, others to patch them up with further absurdities. And the contradictions between dogma and action—e.g. preaching Christian love and waging war in the name of it—could always be solved, or at least absolved, by penance, fines, or service of the church's worldly interests (Tuchman, 1978).

Another recent finding of cognitive psychology concerns multiple reasoning in educated adults. It is the hypothesis that human thinking, far from being sequential, is multiple, in that it consists in a number of parallel threads grouped around a main sequence, which is the one within the conscious area: "Awake or asleep, a number of more or less independent trains of thought usually coexist. Ordinarily, however, there is a 'main sequence' in progress, dealing with some particular material in step-by-step fashion. The main sequence corresponds to the ordinary course of consciousness. It may or may not be directly influenced by the other

processes going on simultaneously" (Neisser, 1963). Introspection confirms this conjecture.

There is a century-old debate on primitive mentality. The early anthropologists noted the great disparities between the savage mind and ours: the poverty of its conceptual stock, its difficulty in generalizing, its incapacity to entertain hypotheses, the clumsiness of its reasoning, and the general lack of interest in abstraction. Being often adjuncts to imperialist expeditions or colonial administrations, some of those anthropologists tended to attribute such backwardness to an inherent biological inferiority. The early liberal reaction to this racist claim was to deny that there is such a thing as the savage mind, and to hold instead that there is only less information: reason is universal, knowledge is not.

Later studies have shown that each of the two contenders held a grain of truth: there is primitive mentality but it has to do with schooling not race. The primitive tribesman cannot perform complicated conceptual operations because he has a poor conceptual kit and does not handle pencil and paper—but he can learn. In particular, knowing only a few numbers, he cannot keep track of any extensive stock, let alone perform calculations, e.g. of distances or speeds—but he can learn.

The anthropologist Sylvia Scribner (1977, p. 486) proposes the following generalizations on the basis of studies concerning proficiency in solving logic problems. "1. *In all cultures*, populations designated as 'traditional' or 'nonliterate' have just somewhat better than a chance solution rate across all types of problem material. [. . .] 2. *Within each culture* there is a large discrepancy of performance between schooled and nonschooled. The major jump seems to occur at levels of education as low as two to three years of school. [...] 3. With schooling, there is little *between-culture* variation in performance for the cultures studied. Grade, rather than society, is most determinative of performance". In short, although actual reason is not universal, potential reason is, and it can be actualized by schooling. So, there is a chasm between the savage mind and the civilized mind, but it is being crossed all the time with the spreading of schooling.

Even in civilized societies most adults seem to reason fallaciously most of the time as soon as they go beyond practical matters. Thus in a typical experiment only 9 percent of the subjects concluded correctly that the premises "All P are M" and "All S are M" do not entail anything about the relation between Ss and Ps (Chapman and Chapman, 1959). In another typical experiment, only half of the subjects drew the correct conclusion (namely "C is worse than A") from the premises that B is worse than A and C is worse than B (cf. Mayer, 1977). When faced with a set of data, the

typical undergraduate conconcts a single hypothesis, is reluctant to consider falsifying data, and fails to consider alternative hypotheses (Johnson-Laird and Wason, 1977). Moreover, casual observation shows that resistance to give up a hypothesis in the face of adverse data mounts with the importance of the hypothesis in one's world view.

Typically, the layperson and occasionally also the scientist and the technologist operate with biased samples—usually collections of items that happen to be noticed and remembered—rather than with random samples; they detect, or fail to detect, covariations suggested by prior beliefs; they tend to expect and forecast what they wish to happen; they tend to pooh-pooh exceptions and criticisms; and generally they adopt beliefs on the strength of first impressions, authority, or fashion. (For these and other fallacies see Nisbett and Ross, 1980.)

The conclusion is obvious: logic is extremely artificial. However, it is not much more so than cooking, building, literacy, or reckoning. Right thinking does not come more naturally than reading or plane piloting: it must be learned. Cognitive psychologists have known this for half a century, and they have also known that most college students will consistently incur fallacies of all kinds unless they have taken courses in logic or in higher mathematics. However, knowledge does not go easily across disciplinary barriers, and so the vast majority of college students graduate without having been exposed to either logic or higher mathematics. And once habits of thought have been acquired it is hard to change them. Thus whereas the person who has been trained in mathematics tends to form clear ideas and reason correctly, the person who has been trained in the humanities is often tolerant of inexactness, and the individual who has had a theological or in general ideological training cannot help thinking in opposites and prefers criticism and debate over constructive reasoning.

1.2. *Formal Reasoning*

Spontaneous or natural reasoning is tacit, undisciplined, and often invalid. Formal reasoning is, by definition, explicit, regimented, and valid. It is formal because its validity is independent of the meaning of the premises and the conclusion. It is explicit in that it involves only explicitly stated premises and inference rules. It is regimented in the sense that it is subject to (though not directed by) principles or rules (e.g. those of the predicate calculus). And it is valid in that, when its premises happen to be true, its conclusions are true as well. (A rule of inference can be construed as a truth preserving function from n-tuples of propositions to propositions.) So far,

only deductive inference has been formalized. There is no reason to hope that other types of inference (e.g. inductive and analogical) will follow suit, for they are content-dependent, i.e. their success depends on the nature of the case.

In ordinary language we often confuse reasons with causes: 'because' can be interpreted either as a causal relation (by cause of) or as a logical relation (by reason of). At first sight this is a confusion: reasons are conceptual objects whereas causes are events. (Bunge, 1959a.) However, on second thought there is something to be said for such conflation: ultimately every reason *is* a brain process and therefore a cause of some other bodily process. So, although not all causes are reasons, every reason is a cause. Still, reasons, to be valid, must comply with certain logical requirements. For example, the truth of A is logically a sufficient reason for upholding "A or B"—but the cause for making the latter claim may be the need or wish to rope B into the argument. Hence, although neurophysiology is ultimately to describe and explain reasoning, it cannot justify it. (Analogy: sociology can explain crime but not condone it.) Only logic can justify reasoning by showing that it fits accepted canons of inference.

The most basic inference pattern is substitution, which is subject to the principle of (invariance under) substitution of variables (terms) or of entire propositions. If in any formula $F(\ldots x \ldots)$ the variable y is substituted for x, then that formula implies $F(\ldots y \ldots)$. For example, $x^2 + x = 0$ and $x = y - 1$ jointly imply $y(y-1) = 0$. The most familiar method for solving systems of linear equations, and the most popular method for evaluating integrals, use the rule of substitution.

The simplest of all the rules of inference involving two premises is of course the *modus ponens*: B is an immediate consequence of A and $A \Rightarrow B$. Combining this rule with the rule of substitution (exchanging A for $\neg B$ and B for $\neg A$) yields the *modus tollens*: $\neg B$ and $A \Rightarrow B$ jointly entail $\neg A$. Many a scientific paper violates either rule. Indeed, a rather common fallacy is this: $A \Rightarrow B$ and B, hence A (or $A \Rightarrow B$ and $\neg A$, hence $\neg B$). This error is in fact committed when a hypothesis is declared true because some of its consequences have been confirmed.

Needless to say, the detectives in good mystery stories do not commit such fallacies. A typical example is this. After some inquiries the inspector gathers all the people involved in the crime and presents them his premises:

1. Every criminal has a motive and an opportunity. (Law of criminology.)
2. The butler had both a motive and an opportunity. (Empirical datum.)

3. Nobody except the butler had both motive and opportunity. (Plausible hypothesis.)

All three premises are necessary and sufficient to draw the conclusion that the butler did it. This proof can be obtained by *reductio ad absurdum*: Suppose someone else, call him or her C, did it. Then C had both motive and opportunity. But this contradicts premise No. 3. Hence C is identical with the butler. The argument is an example of valid *hypothetical* reasoning, for it involves two hypotheses going beyond the hard datum 2, namely 1 and 3.

To be sure, we are not always given conditionals: sometimes our task is to prove them. In mathematics, to prove that A implies B one can often assume A and check whether it entails B (i.e. whether "$A \Rightarrow B$" is a tautology). For, if it does, then a fortiori $A \Rightarrow B$. (A second method is to assume $\neg B$ and find out what this entails concerning A. If it entails that $\neg A$, the conditional has been established.) In factual science and technology the method for establishing $A \Rightarrow B$ is quite different. One enforces A and checks whether B obtains. If it does, the conditional is confirmed but not proved, for it may well happen that only the favorable cases were observed during the given experimental run. If B does not obtain, the conditional has been refuted—or so it seems. In general, there are no rigorous proofs of factual statements.

Strictly speaking, mathematical proofs are not absolutely rigorous either, and this simply because there are no absolute (definitive) standards of rigor. Given any proof, it is likely that it can be improved on, if not right now, then in a few years or centuries. Moreover, logical standards are changeable. The history of the calculus is littered with proofs that were regarded perfect at one time but turned out to be defective in some respect or other. However, there is a big difference between mathematical proof and factual confirmation: whereas the former can *in principle* be rendered unassailable relative to the prevailing standards of rigor, factual confirmation remains in *principle* assailable, not only because its premises are uncertain but also because it involves plausible rather than demonstrative inferences. Caution: the fallibility of a mathematical proof and of a factual confirmation does not entail that we must ultimately resort to some extrinsic sanction such as consensus. It only suggests that further research may be needed.

Our discussion of deductive inference has not depended upon the notion of truth. To be sure, we did characterize a rule of deductive inference as a truth-preserving function from n-tuples of propositions to propositions. However, this condition is needed to check the validity of any proposed

rule of inference, not to perform any deductions. (Thus the obviously false premises "All men are immortal" and "Donald Duck is a man" entail rigorously "Donald Duck is immortal".) It is just as well that deducibility should not involve truth, for often (*a*) we do not know whether our premises are true (but are just trying them out), and (*b*) our premises are at best half-true.

Logic, the theory of deduction, is concerned to save consistency, not to guarantee truth. We want consistency for the following reasons: (*a*) a contradiction entails any proposition, and is thus a source of unwanted proliferation (conceptual cancer); (*b*) consistency is necessary (though not sufficient) for a set of propositions to form a system allowing one to jump safely from one component to another. Logic (or rather metalogic) can diagnose consistency and inconsistency but is impotent to advise what to do in the face of inconsistency. To be sure, if a set of premises turns out to be inconsistent then we can strike off some premises to obtain one or more contradiction-free subsets of premises, but logic will not tell us which, if any, such subsets is formed by true premises.

For example, the set of premises $\{A, B, A \Rightarrow \neg B\}$ is inconsistent. In fact, the 1st and 3rd imply $\neg B$, which contradicts the 2nd; and the 2nd and 3rd imply $\neg A$, which contradicts the first. The three minimal surgeries yielding consistent sets of premises are:

(i) Drop A, leaving $\{B, A \Rightarrow \neg B\}$, which reduces to $\{\neg A, B\}$.
(ii) Drop B, leaving $\{A, A \Rightarrow \neg B\}$, which reduces to $\{A, \neg B\}$.
(iii) Drop $A \Rightarrow \neg B$, leaving $\{A, B\}$.

In the case of mathematics all three subsets may be correct though of unequal or even null interest. If we are to choose among them it will be on grounds of richness or beauty, economy or applicability. If the premises are factual, compatibility with the available evidence and with other theories may indicate which of the three subsets, if any, is the best (truest and deepest) set. But it may well happen that all three are inadequate—or that the evidence is so hazy that neither is completely ruled out by it. In conclusion, *pace* Rescher (1976), there is no a priori general theory of plausibility that could help us choose among the above contradiction-free subsets.

Consistency is an absolute desideratum relative to each particular body of knowledge, in particular theory. It is not equally desirable with regard to the totality of knowledge at a given moment: if overall consistency were required at every moment, then no rival views would be tolerated, and

consequently we would be unable to correct any errors. So, inconsistency is a fact of life and we must learn to eliminate it (by uprooting errors) rather than to avoid it.

Dialectical philosophers hold that contradiction is not only unavoidable but also desirable to keep things moving, for contradiction among ideas (and even among facts) is the source of all change. The evidence goes against the universality attributed to the genetic power of contradiction. Indeed the history of ideas shows that, whereas some conceptual contradictions have indeed promoted progress (by calling for strenuous efforts to surmount them), others have resulted in stagnation or even regress. Here is a random sample of counterexamples to the dialectical thesis. (*a*) The contradiction, within the Pythagorean school, between its tenets concerning rational numbers and the discovery of irrationals (such as the square root of 2), brought the efforts of the members of the school to a halt. (*b*) The contradiction between the subjectivist and the frequency interpretations of probability has produced no new results except for confusion. (*c*) The contradiction between the wave and the particle interpretations of the quantum theory has held back work in the foundations of the latter. (*d*) The contradiction between classical mentalistic psychology and behaviorism kept psychology from making any breakthroughs for decades. (*e*) The contradiction between individualism and holism in social science is still bogging it down. The moral is clear: Contradictions, when protracted, are unhealthy because they generate sterile and acrimonious debates rather than new ideas or new experiments. Contradiction does not generate anything: at most it may motivate some people to overcome it.

We often need to estimate the degree of agreement or disagreement between two sets of propositions, e.g. between two scientific theories, or two ideologies, or a scientific theory and an ideology. Actually we need two different measures, one for the degree of agreement between two sets of propositions, and another for the degree of compatibility. The two measures must be radically different because two sets of propositions may be quite different yet compatible (e.g. an atomic theory and a sociological theory), or similar yet mutually incompatible (e.g. two rival theories).

(Consider two sets of propositions $P = \{P_i | 1 \leq i \leq p\}$ and $Q = \{q_j | 1 \leq j \leq q\}$. We define the *degree of agreement* between P and Q as the fraction of the number of propositions they share, i.e.

$$A(P, Q) = (1/p + q) \cdot |P \cap Q|.$$

The degree of *disagreement* between P and Q is defined as

$D(P, Q) = 1 - A(P, Q)$. The values of A and D are comprised between 0 and 1. Such values can be computed only for finite sets of propositions. In the case of theories we may take the sets of their postulates; in the case of rival theories it pays to enrich the postulates with all the standard theorems. Two different presentations or organizations of the same theoretical material are supposed to agree fully with one another, i.e. $A(P, Q) = 1$. On the other hand two genuinely rival theories must share some but not all their propositions, i.e. $0 < A(P, Q) < 1$.)

(Let us now introduce the concept of degree of compatibility between two sets P and Q of propositions. They will be totally incompatible if every member of P contradicts each component of Q—an extremely unlikely case. We abbreviate 'p_i is incompatible with q_j' to '$p_i | q_j$', and correspondingly 'p_i is compatible with q_j' to '$\neg(p_i | q_j)$'. We now define the *degree of compatibility* between P and Q as the fraction of the number of pairs of compatible propositions:

$$C(P, Q) = (1/pq)|\{\langle p_i, q_j\rangle \in P \times Q | \neg(p_i | q_j)\}|.$$

The dual is of course the degree of *incompatibility* between P and Q, namely $I(P, Q) = 1 - C(P, Q)$. The values of the two measures are comprised between 0 and 1. If $C(P, Q) = 1$, P and Q are fully compatible, even if they do not agree at all, whereas they are totally incompatible if $C(P, Q) = 0$. Any value between these extremes indicates incomplete compatibility.)

Can logic help us obtain new knowledge? Nearly all of the 17th and 18th century philosophers derided logic for its alleged impotence to do so. They demanded instead that a logic of invention and discovery be instituted. Although such a logic was never born, it was much praised. And in the meantime the contempt for logic blocked its advancement. (You do not farm well a land that you despise or hate.) Logical investigations were resumed two centuries later by mathematicians, but the calumny that logic is useless to advance knowledge still remains. Even after mathematical logic was well established, a distinguished philosopher such as Schlick (1925) ignored its role in research and suggested that, since science studies only facts, and there are no negative facts, it should make do with a negation-free logic. This is a correct inference from the thesis that scientific research is just the collecting of trivial facts. It falls down as soon as one realizes that there is no science without conjecturing and arguing, which involves counterconjecturing and counterarguing. Certainly there are no negative facts, but negation (a "positive" brain process) is part of every inquiry process. Statements of the form "A is not B" are just as important

as those of the form "A is B"; i.e. identity and inclusion are not more important than difference and exclusion.

When a mathematician, a theoretical scientist or a technologist prove a new theorem, they produce genuine new knowledge with the help of the logical theory underlying (presupposed by) the theory or theories they employ. (And, since the most common types of proof are by *reductio ad absurdum* and by counterexample, negation is usually involved in such deductive processes.) To be sure, logic does not supply any substantive premises: it acts as midwife rather than parent. Nevertheless its intervention is necessary to produce or control the new results. If deduction did not yield new knowledge we would not care to perform it and, in particular, we would not care to build or use theories proper, all of which are hypothetical-deductive systems.

It is often rejoined that the conclusion of a deductive inference is "contained" or "included" in the premises, so it is not actually new. This is sophistry. In the first place not all the conclusions of a deductive inference occur among the premises. Thus "All A are C" follows rigorously from "All A and B" and "All B are C" (as well as from "All A are X" and "All X are C"), but is not contained in these premises. Likewise one gains new knowledge, by deduction, from the data that $a > b$ and $b > c$, namely $a > c$.

Secondly, even if all deduction were merely unpacking, it still would supply new knowledge: if nobody knew the conclusion before it was drawn, then the conclusion was not an item of knowledge—it only became one. What is true is that we may *feign* that the conclusion exists (formally, not epistemically) the moment the premises (and the underlying logic) are laid down. Likewise the value of a continuous function at a given point is "given" by a certain formula containing the values of the function and all its derivatives at some other point, but the former is unknown as long as the computations indicated by the formula have not been performed.

In short, deduction is a prime tool of inquiry, for it allows us to explore the (a priori unknown) consequences of our assumptions and definitions. And, because deductive reasoning is formalizable and subject to explicit rules, one is justified in calling *hypothetico-deductive method* the pattern of reasoning consisting in assuming, deducing, and checking. No other type of inference is formalizable and subject to explicit rules, and therefore capable of becoming the nucleus of a method. In particular, although we all use plenty of intuition, analogizing and induction, there is no such thing as the *method* of intuition, or analogy, or induction. There are only philosophical tracts on such ghostly subjects.

Finally, are there things we cannot reason about? All we know is that, every so often, we come up against certain limits to knowledge that are subsequently overcome. Thus the Eleatic philosophers believed change to be contradictory and therefore contrary to reason; and, being rationalists, they declared change to be illusory. Thus a moving arrow was said *to be* and *not to be* at a given place at a given time. Hegel kept this archaic mode of thought and admitted the Eleatic thesis that change is contradictory, but claimed that this only calls for the dialectical mode of thought. Bergson chose the other horn of the ancient dilemma: he claimed that only intuition can grasp change—whereas science, which is rational, cannot. We see no contradiction in change. Instead of saying that the arrow is, and is not, at a given place at a given time, we say that it moves with such and such instantaneous velocity through a given place. We replace being with becoming as the basic category, and use the differential calculus to reason about change. One man's limits to reason is another's challenge to improve reasoning.

2. Plausible reasoning

2.1. *Analogical Inference*

The most basic judgment we can form about a collection of items is that they are, or fail to be, similar in some respect(s)—i.e. that they are analogous or disanalogous. Judgments of analogy underlie a number of judgments of other types, in particular collecting or grouping (Ch. 5, Section 1.2) and inductive generalization (Section 2.2). We make them all the time in daily life and in all fields of inquiry. Here is a tiny random sample of analogies that have been fruitful at one time although they ended up by blocking further research: set–collection of things, gas–solute, electric current–fluid current, light–sound, economic competition–natural selection, optics–mechanics, and rumor propagation–epidemics. We tend to remember only the fruitful analogies. The list of misleading analogies, if it could be drawn, is likely to be more impressive.

Analogical inference is based on analogical judgment but goes beyond the latter: from the fact or assumption that two objects are analogous in some known respect(s) it "concludes" that they *may* be analogous in some other unknown respect(s) as well. Thus organicists, likening society to the human body, conclude that the workers, being the muscles of society, ought to obey the brain of society, namely its ruling class. And Copernicus,

likening the solar system to a monarchy, argued that, just as the monarch is surrounded by his vassals and courtiers, so the sun must be at the center of the solar system. Both are of course implausible reasonings, not plausible ones. They are faulty because they rest on wrong analogy judgments: society is not like an organism, and the solar system is not like a monarchy. On the other hand Aristotle's classical example of analogical reasoning, namely "Gills are to water what lungs are to air", was plausible and fruitful. It was so because it rested on two genuine analogies: gill—lung and water—air. A necessary condition for a plausible and fruitful analogical inference is then that it rests on genuine analogies. We should therefore elucidate the rather vague notion of analogy.

There are several ways of exactifying the notion of analogy, and each is suitable for a given set of problems. One of them involves the modeling of the analogous objects as sets. In this case one can elucidate a number of analogy strengths, from mere some–some correspondence to homomorphism, to isomorphism or perfect similarity. (Cf. Bunge, 1973b, Ch. 6.) A second method is to count the number of properties shared by the analogous things: see Vol. 3, Ch. 2. A third applies when the things concerned are systems, i.e. objects with a definite composition, environment, and structure. We proceed to elucidate this notion (Bunge, 1981b).

Consider two systems, each modeled qualitatively by a triple *composition–environment–structure* (Vol. 4, Ch. 1). Each component of this triple is a set. In comparing the triples representing two systems we compare sets (of components, environmental items, or relations). Therefore we must start with the concept of similarity or analogy of two sets. We stipulate that two sets are similar with respect to a third set if they overlap with the latter. More precisely, if A, B and C are nonempty sets, then *A is similar to B with respect to C* if, and only if, $A \cap C \neq \varnothing$ and $B \cap C \neq \varnothing$. We can now define the concept we want:

Let s_1 and s_2 be two concrete systems with known compositions, environments, and structures. Then:

(i) s_1 and s_2 are *substantially analogous* iff their compositions are similar, i.e. if there is a natural class or species that has a nonempty overlap with those compositions;

(ii) s_1 and s_2 are *environmentally analogous* iff their environments are similar, i.e. if there is a natural class or species that has a nonempty overlap with those environments;

(iii) s_1 and s_2 are *structurally* (or *formally*) *analogous* iff their structures are similar, i.e. if there is a family of relations that has a nonempty overlap

with those structures. (In particular, two systems are *nomologically analogous* if they share some laws.)

All organisms are substantially analogous, if only because they all contain DNA molecules. Fish and marine mammals are environmentally analogous. The human hand and the bat wing are structurally analogous (homologous).

The above concepts may be quantitated with the help of the following concept of degree of similarity between sets. If A and B are sets intersecting a third set C, then A and B can be said to be fully analogous with respect to C if both have the same number of elements of kind C, i.e. if the cardinalities of $A \cap C$ and $B \cap C$ are the same. The degree of dissimilarity between A and B with respect to C is then measured by the difference between $|A \cap C|$ and $|B \cap C|$. More precisely, we stipulate that

$$d_C(A, B) = \frac{||A \cap C| - |B \cap C||}{\max\{|A \cap C|, |B \cap C|\}}.$$

Therefore the *degree of similarity between A and B with respect to C* may be taken to equal the complement of $d_C(A, B)$ to unity, i.e. $s_C(A, B) = 1 - d_C(A, B)$. The values of the degree of relative similarity are comprised between 0 and 1. If A and B contain no C's, then $s_C(A, B) = 0$. But if there are just as many C's in A as in B, then $s_C(A, B) = 1$ even though they are not the same individuals. For two sets to be similar in the given respect C they just have to contain at least one member of kind C.)

By applying this concept of similarity between sets to our previous notions of systems analogy we obtain the concepts of degree of substantial, environmental, and structural analogy. Moreover, by adding all three and dividing by 3 we get the concept of *degree of total analogy*. Finally, we can invert the process to obtain the qualitative concepts of analogy in terms of the quantitative ones: Two systems are (*a*) *analogous* iff their degree of total analogy is greater than 0; (*b*) *weakly analogous* iff their degree of total analogy, though nonvanishing, is close to 0; (*c*) *strongly analogous* iff their degree of total analogy is close to 1.

A proposition of the form "A and B are analogous" is *true* if, in fact, the degree of analogy between A and B is greater than 0. However, truth is not enough, for it may be deep or shallow, fruitful or barren. The judgment of analogy will be *superficial* if the systems in question are weakly analogous. (*Example*: the brain–computer analogy.) The judgment will be *deep* if the systems in question are strongly analogous. (*Example*: the man–ape analogy.) However, even deep analogies may not be fruitful. The fruitful-

ness of an analogy depends not only upon its degree but also on the problem at hand and the brain that uses it. Many people had noted the analogy between man and other primates; in the Middle Ages it was a common belief that monkey was a fallen man. However, it took a Charles Darwin to use this analogy to place man and ape in an evolutionary perspective.

Because the fruitfulness of an analogical inference depends on the actual degree of analogy as well as on the imagination of the user, there can hardly be a general theory of analogical reasoning. The most we can do is to formulate some useful methodological (not logical) rules, such as:

R1. Use only deep analogies—i.e. make sure that the degree of (substantial, environmental, structural, and if possible total) analogy is significantly greater than 0.

R2. Regard all analogies as heuristic devices that may have to be discarded eventually.

The first rule is obvious: weak analogies lead nowhere or, worse, they may block inquiry. For example, the analogies between brain and hydraulic system, electrical network, telephone exchange, or computer, are not only extremely weak: they have also misled uncounted workers in neuroscience and in psychology into studying technological gadgets rather than living brains. (More in Vol. 6.)

The second rule is justified as well: even if initially fruitful, an analogy is no substitute for an investigation of the peculiarities of the thing of interest. Thus the quantum-classical analogies were initially useful as rough guides, but ended up by obscuring the originality of the quantum theories as well as the nonclassical nature of their referents (Bunge, 1973a). And the information-theoretic metaphors employed in biochemistry, molecular biology, and genetics, helped crack the genetic "code" but they are now blocking the way to an understanding of the mechanism of protein synthesis and, in general, the way "genetic information" is locked in nucleic acids. (More in Vol. 6.)

Because rigorous reasoning is nonanalogical, scientific papers in advanced science and technology are not supposed to resort to analogical reasoning—not even when proposing analogies. If they do, they do so inadvertently. On the other hand analogy and analogical inference are rampant in the popularization literature, for they are (wrongly) assumed to convey technical ideas to the lay. (Actually they skirt the problem of conveying them.) And, since most philosophers have no access to the original scientific or technical literature, they sometimes get the impression

that analogy, far from being just a heuristic trick and a pedagogical prop, is of the essence of science and technology. (For the quaint thesis that scientific theories are analogies or metaphors, see Black (1962) and Hesse (1966).) Occasionally there are also more serious motivations for this view, or analogism. One is fictionism, or the doctrine that we cannot know what and how things really *are* but at most what they are *like*, so we should present our hypotheses and theories as stating only that things behave or appear *as if* they were such and such. (Fictionism has its roots in Kant, was sketched by Nietzsche, and worked out by Vaihinger (1920).) Another root of analogism is the attempt to present science as no better than religion: just as (according to Aquinas) we can know God only by analogy with worldly entities, in particular ourselves, so science would know the world only by analogy with our experience.

In short, analogies and analogical inferences are indispensable tools of inquiry. But analogies can be superficial or deep, and analogical inferences can be misleading or fruitful. It is therefore necessary to assess the respect and degree in which an analogy holds, and to check the conclusion of an analogical argument. And we must not conclude to analogism from the fact that analogy is pervasive, if only because it is not more so than other categories.

2.2. *Statistical Inference and Hypothetical Reasoning*

Induction is, of course, generalization from known (or partially known) instances. Every higher vertebrate performs inductive inferences, and scientists and technologists are no exception. Many of the generalizations, true and false, we form and use in daily life are inductions. (Perhaps most of them are generated by only a handful of cases.) And every time we evaluate the performance of hypotheses, theories, methods and plans, we formulate inductive judgments. So, induction occurs sometimes in the inception of general constructs and always in their validation.

We distinguish the following types of *inductive inference.*

(i) *First degree induction*: from instances to low level generalization. *Pattern*: "All A's up to the nth were found to be B's. Hence it is likely that all A's are B's." *Example*: All the measured values of the speed of light in a vacuum are the same to within experimental error. Hence it is likely that all light waves in a vacuum have the same speed.

(ii) *Second degree induction*: from low level generalizations to higher level generalization. *Pattern*: "Law L holds for every member of a family of sets

up to the nth. Hence it is likely that law L holds for every member of the family." *Example*: The theory of evolution holds for all studied species. Hence it is likely to hold for all biospecies.

(iii) *Statistical specification*: from population to sample. *Pattern*: "The observed fraction of P's in the population U is p, and S is a random sample of U. Hence it is likely that the fraction of P's in S be near p." *Example*: The last census showed that 10 percent of the population are unemployed. Hence it is likely that about 10 percent of the population of Medianville, which is pretty representative, are unemployed.

(iv) *Statistical generalization*: from sample to population. *Pattern*: "S is a random sample of the population U, and the observed fraction of P's in the random sample S of U is p. Hence it is likely that the fraction of P's in U be near p." *Example*: In a random sample of college students amounting to the one-hundredth part of the total student body 1 percent were found to be original thinkers or doers. Hence it is likely that the corresponding percentage in the whole student population be about 1 percent.

Note the following traits common to all of these inference patterns. First, neither is wild: they all proceed from supposedly sound (checked) premises. Second, neither leads to new concepts; in particular, all the theoretical concepts occurring in the conclusions appear also in the premises. Third, neither is conclusive or demonstrative: they are not of the form "A, hence B", but rather "A, hence it is likely that B", with no specification of the degree of likelihood. Thus room is made for failure of the inferential process. For instance, in the example of (iii) Medianville, though representative on the whole, may be particularly lucky, or unlucky, at the moment the inference is drawn—e.g. because it happens to host an industry that has suddenly boomed or slumped. And in the example of (iv) the student sample, though representative on the whole, may be exceptionally rich or poor in talented students at the time the sample is drawn—e.g. because talented students tend to concentrate in certain colleges.

Consequently neither of the above inference patterns can be erected into a rule or principle, e.g. the "principle of induction", more often praised or criticized than formulated. The conclusions of all such plausible reasonings are just as hypothetical as their premises. Consequently they call for further testing. Besides, neither the premises nor the conclusions occurring in inductive inference are enough: we also need higher level hypotheses, which cannot be generated by studying particulars, if only because they contain concepts that are far removed from observation. The task of induction is not to generate such high level hypotheses in step by step

processes, but to take part in their empirical tests, as will be seen in Ch. 12.

In addition to the above four patterns of inductive inference we shall mention four kinds of *hypothetical reasoning*.

(i) *Weak modus ponens*: weak detachment of consequent upon weak assertion of conditional or of antecedent. *Patterns*: (*a*) "If p then q, and 'p' is verisimilar. Hence 'q' is verisimilar"; (*b*) "'If p then q' is verisimilar, and p. Hence 'q' is verisimilar." (*c*) "'If p then q' is verisimilar, and 'p' is verisimilar. Hence 'q' is verisimilar." *Example*: Heavy smoking is likely to cause lung cancer. John is a heavy smoker. Hence John is likely to catch lung cancer."

(ii) *Weak modus tollens*: weak rejection of antecedent upon strong or weak assertion of conditional and weak or strong negation of consequent. *Patterns*: (*a*) "If p then q, and '$\neg q$' is verisimilar. Hence '$\neg p$' is verisimilar". (*b*) "'If p then q' is verisimilar, and $\neg q$. Hence, '$\neg p$' is verisimilar." (*c*) "'*If* p then q' is verisimilar, and '$\neg q$' is verisimilar. Hence '$\neg p$' is verisimilar." *Example*: It is likely that philosophy flourishes only in freedom. Now, there is no freedom in Dogmaland. Hence it is unlikely that philosophy flourishes in Dogmaland.

(iii) *Strong reduction*: weak assertion of antecedent upon assertion of consequent. *Pattern*: "If p then q, and q. Hence maybe 'p' is true." *Example*: brain injuries cause intellectual deficits. Now, Peter has recently converted to a mystical doctrine. Hence it is possible that Peter suffered a brain injury.

(iv) *Weak reduction*: weak assertion of antecedent upon assertion of consequent and strong or weak assertion of conditional. *Patterns*: (*a*) "If p then q, and 'q' is verisimilar. Hence it is possible that 'p' is true." (*b*) "'If p then q' is verisimilar, and q. Hence it is possible that 'p' is true." (*c*) "'If p then q' is verisimilar, and 'q' is verisimilar. Hence it is possible that 'p' is true." *Example*: Famines are likely to cause political unrest. Now, there was a revolution in A during period B. Hence it is possible that there was a famine in A during B.

Like inductive reasoning, plausible hypothetical reasoning is educated rather than wild, as well as inconclusive: its conclusion is at best another conjecture worth being investigated. But, unlike inductive reasoning, plausible hypothetical reasoning may proceed from premises containing transempirical concepts. For this reason it occurs not only in the early stages of empirical research and in the testing stage but in all stages of scientific and technological inquiry.

Inferring from symptoms, or indicators, is a common species of

hypothetical reasoning and, in particular, of weak reduction. Thus since all infections cause an abnormal count of white blood corpuscles, we tend to attribute the latter to some infection; however, it may have a different cause, such as cancer of the bone marrow. Likewise we tend to attribute criminality and drug addiction to poverty, but some times it is caused by alienation.

Most indicators in the life sciences and in the social sciences are ambiguous, i.e. they occur in hypotheses of the form *If A or B then C*, where 'C' denotes the observed variable adopted as an indicator of A or B. Therefore it is advisable to employ a whole battery of indicators in order to narrow down the set of possible causes. But the only effective way of removing the uncertainties in inferences from symptoms is to employ functional relationships between symptoms and causes. Thus the barometer gives an unambiguous reading of atmospheric pressure because it is built on the law of the linear relation between atmospheric pressure and barometer height; and it can be used to give an unambiguous value of the altitude on the strength of the exponential law relating altitude and atmospheric pressure.

Plausible reasoning, whether analogical, inductive, hypothetical, or of some other kind, cannot be regimented any more than intuition or insight can. First, because it depends essentially on the nature of the case. Second, because it depends on the talent of the subject to notice or conjecture relevant features or patterns. Third, because one and the same set of data is compatible with any number of alternative hypotheses: Figure 6.1. In sum there can be no *logic* of plausible reasoning.

However, plausible reasoning can be controlled. Actually it is being controlled by mathematical statistics, not by philosophy. Indeed, the tasks

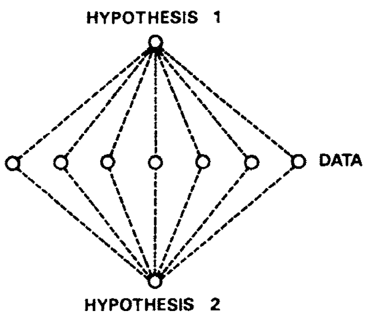


Fig. 6.1. At least two different (rival) hypotheses are compatible with any given set of data.

of that discipline are not only to help collect, screen and organize data, but also to help compare the latter with hypotheses in order to evaluate both (usually the hypotheses). But, of course, mathematical statistics produces neither raw data nor the hypotheses accounting for them: it is a device for control not generation. For example, it tells us that statistical generalization and specification are plausible only if the sample is random and its size not too small: otherwise violent fluctuations around the frequency will be observed. (Sample size, though essential, is less so than randomness.)

We all pay lip service to mathematical statistics but few of us succeed in avoiding committing statistical fallacies such as the gambler's fallacy. This is the belief that every segment of a random sequence (or every random sample of a population) must have the same properties as the entire sequence (or population). Thus if in coin tossing the inexperienced gambler gets five heads in a row, he may be inclined to bet that the next flipping will produce a tail: he believes that there must exist a force producing a deviation that will cancel the previous deviation from the long run 0.5 frequency. This fallacy is committed not only by gamblers but also by a disquieting large fraction of scientific researchers (particularly in psychology and social science); moreover, they exhibit undue confidence in early trends, overestimate significance, and offer causal explanations for any local discrepancies instead of paying attention to sample variability (Tversky and Kahneman, 1971). This finding shows how inadequate our statistical education is, and it suggests that the teaching of probability and statistics should start earlier than usual, perhaps in primary school.

3. Discussing

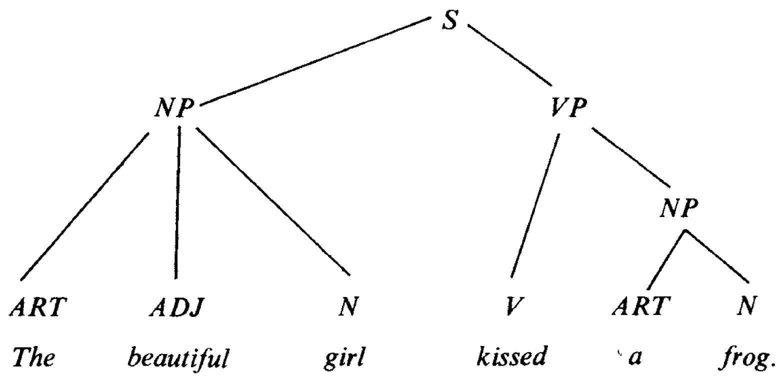
3.1. *Analysis and Criticism*

Can we learn from discussing ideas, e.g. from analyzing them and arguing about them? Rationalists reply in the affirmative, and some of them claim that analysis, criticism and argument are the very best cognitive methods. The schoolmen of all times assent because they hope to extract all the knowledge locked in allegedly canonical and infallible texts. The linguistic philosophers, because they regard ordinary language as the repository of wisdom. The dialectical philosophers, because they believe that new ideas can be generated only by the clash of old ideas (or the clash between ideas and facts). And the critical rationalists, because they are interested

primarily in the weeding out of false ideas. On the other hand, empiricists deny all this: to them only experience, in particular observation, can give us new knowledge, whereas discussion is unprofitable and perhaps even (as Comte said) frivolous. Who is right? Once again it will turn out that each of the contending parties holds a grain of truth.

To analyze a complex object, such as a system, is to exhibit its components, environment (or context), and structure (or organization). The analysis can be partial (decomposition into subsystems) or total (decomposition into ultimate or elementary components). It can be empirical (e.g. chemical) or conceptual (e.g. logical). In every case the components, the (relevant) environmental items, and the relations or links holding the system together are identified and classed. A few examples will show the gist and scope of such procedure.

Example 1. Structural analysis of a sentence. Look at the tree diagram representing the surface structure of the English sentence *The beautiful girl kissed a frog.*



The species or lexical categories are ART (icle), ADJ(ective) and N(oun); the genera NP (noun phrase) and VP (verb phrase). One and the same object, namely the sentence S, is conceived of as a system and analyzed on two levels: subsystems (made up of objects on the genus level, NP or VP), and elementary objects (belonging to species such as N and V). The analysis uses then a word systematics that is taken for granted, and which differs somewhat from language to language. (Thus Latin lacks the ART species.)

Example 2. Logical analysis of the proposition expressed by the sentence in Example 1. A possible analysis is this:

For some x, x is a frog and b kissed x,

where

$b =$ The y such that y is a girl and y is beautiful at fixed place p and time t.

Example 3. Mathematical analysis of proposition in Example 2. Every one of the logical categories occurring in the given proposition can be analyzed further, i.e. into deeper concepts. For example, the predicate "frog" can be construed as a function from the set of all reptiles (or all vertebrates, or all animals, or all organisms) to the set of all propositions contining that predicate. Likewise the predicate "beautiful" can be analyzed as a function from the cartesian product of the set of all objects, the set of all places, and the set of all times, to the set of all propositions containing that predicate. The connective "and" can be analyzed as a function from pairs of propositions to propositions. The particularizer "Some" can be analyzed as "Not-for-all-not", where "all" and "not" are primitives, or in a more sophisticated way. Likewise the descriptor "the y such that" can be analyzed *à la* Russell or as in Vol. 2, Ch. 9.

In all three cases the "environment" (context) of the system being analyzed has been taken for granted: the English language in the first, the predicate calculus in the second, and mathematics in the third. And every analysis has presupposed some body of knowledge about such contexts—in particular a knowledge of basic categories such as "noun", "predicate", and "function". The logical analysis (Example 2) has also presupposed a bit of factual knowledge, which on closer examination turns out to be dubious. In fact, that analysis presupposes that beauty is absolute (independent of the beholder); it also presupposes that place and time are absolute (independent of the reference frame). If relativized, the predicate turns out to be of degree 5 or higher, not 3.

In performing each of the three analyses we had to invest a certain amount of knowledge. But we gained some new knowledge too. For, even supposing that we knew the components of the system before analyzing it (since we fashioned it ourselves), we did not know the kind or type to which each of them belonged, or how they were related to one another. To be sure, the output has been puny by comparison with the input. However, the same can be said of all kinds of analysis. Think of the analysis of a function as an infinite series, or the analysis of a chemical compound into atoms, or the analysis of a society into economic, political, and cultural subsystems. In all these cases the knowledge investment (data and general principles) is far greater than the knowledge produced. But what is important is that *(a)*

analysis *can* produce new knowledge—empiricism notwithstanding, and (*b*) analysis is only *one* of the ways of gaining new knowledge from available knowledge—rationalism notwithstanding.

Analysis can be valuable, but not all analysis is. Thus linguistic analysis, as practised by Wittgenstein and his followers, and hermeneutic analysis, the daily fare of lawyers and popular among theologians, political ideologists, and contemporary German philosophers, are ambivalent. The former because it is shallow and sometimes misleading: it makes no use of logic—let alone mathematics or science—and is limited to examining the use of ordinary language locutions. Its results are usually very modest, as must be expected from a "philosophy without tears" (Russell). And sometimes they are grotesque, as when Ryle (1954) concluded, from an examination of what he called 'the grammar of the word *perceiving*', that seeing and hearing are not states or processes, hence not something to be located, inspected and measured. Correct conceptual analysis calls for more than a mastering of one's dialect: it requires substantive knowledge and fine analytic tools, often mathematical ones.

As for hermeneutics, or the interpretation of texts, it has three uses. One is to arrive at an understanding of the original text, another is to evaluate current actions or ideas in the light of some code, and a third is to help us understand or modify current events. In the first case, the hermeneutic analyst attempts to explain the given text in contemporary terms and to place it in its original cultural context. This enterprise is as risky as it is useful, for it forces the authors of the past to think and write like ourselves.

A second use of hermeneutic analysis is to examine certain actions or ideas to find out whether they conform to some doctrine or code. It is standard procedure in law. The basic inference pattern is this: *A holds or does B. Now, B is compatible (or incompatible) with canonical text C. Hence A is right (or wrong) on C.* The doctrine or code itself is not questioned. So, the analysis is useful if the canonical text happens to be correct or just, but it can be noxious if the text is wrong or unjust. In any case this kind of hermeneutic analysis does supply some knowledge—e.g. that A did say or do something fitting (or clashing with) a certain canonical text. Moreover, if what A said or did is judged as correct or just on a different standard, then that piece of hermeneutic analysis will suggest that the canonical text is wrong on the given point, which in turn may elicit its reform.

Finally, a third kind of hermeneutic analysis is to try and understand what happens to and around us in terms of some text written hundreds or

even thousands of years ago for different people in different circumstances. Such analysis is worthless or worse, as it involves a distortion of the text, the contemporary events, or both—for otherwise the events could not possibly fit the text. Needless to say, this use of hermeneutic analysis is common among ideologists wishing to confirm their beliefs or to use them to praise or condemn some of their contemporaries. Arguing with analysts of this kind is unlikely to produce new knowledge, for their intention is apologetic, not inquiring.

Conceptual analysis can be regarded as a means for uncovering composition, environment or structure, or as part of criticism. And criticism is of two kinds: *apologetic* (or in defense of a belief system) or *inquiring* (or aiming at spotting error). In turn, inquiring criticism can be *negative* or *constructive*. The former stops when error has been found and exposed as such. Constructive criticism goes farther, to error correction. The proofreader, the dedicated teacher and the productive investigator, technologist, and manager are constructive critics. On the other hand the literary critic and the art critic, as well as the typical book reviewer, are negative critics. Negative criticism is easier and therefore more common than constructive criticism: (*a*) it bears on something that has already been set up (by oneself or someone else), and (*b*) the denial of a proposition has better chances of being true than its assertion. (*Example*: if one states that the current world population is exactly 4 billion, he shall be in error. The denial of that statement is true—a cheap truth, for it gives no positive information.)

Criticism is necessary in all fields of inquiry because an aim of all inquiry is truth, and no research is free from error. We are bound to make some erroneous assumptions, observations, computations, or inferences; we are bound to employ vague concepts or inadequate methods; and we are bound to disregard some variables and overrate others. Any such error may spoil a research project if it goes undetected and uncorrected, but none of them counts in the end if spotted and repaired in time. This is why criticism is tolerated, nay encouraged, in science, technology, the humanities, and enlightened management and politics. For the same reason criticism is discouraged elsewhere, particularly in institutions devoted to the maintenance of dogma, authority, or privilege. Thus churches meet criticism with excommunication, the armed forces with court martialling, and some political parties with expulsion.

The epistemology of science and technology involves methodological (though not systematic) skepticism, a critical attitude that spares nothing

excepts the ideals of objectivity and rationality. Everything, from theories to data, from methods to institutions, can become the object of scientific doubt or technological redesign. Science and technology have institutionalized skepticism: "Organized skepticism involves a latent questioning of certain bases of established routine, authority, vested procedures, and the realm of the 'sacred' generally" (Merton, 1973, p. 264). For this reason science and technology are bound to clash sooner or later with all other institutions, for these usually demand faith and loyalty.

The critical attitude, essential in science, technology and the humanities, seems to be rather uncommon elsewhere, and tolerance of criticism quite recent. There must have been dissenters in all human societies, but they are likely to have been persuaded or suppressed by persons wielding authority. All societies have belief-enforcing mechanisms, and only a few tolerate open criticism, particularly of basic social policies. However, in most cases no repression is required to exact conformity, for most people are ready to believe or do what they are told.

In short, criticism is necessary to clarify ideas and eliminate errors. However, criticism should be neither overdone nor overrated. It should not be overdone because it may kill good ideas in the bud, when they are still hazy and have not shown their worth. This is why scientists and technologists resent hasty criticism: they know that every idea is born incomplete and defective, and it should be allowed to develop before being shot down. And criticism should not be overrated (as Popper (1963) does) as the main moving force of every intellectual development because, before an idea can be analyzed, criticized, or rejected, it must have been thought up by somebody. After all, it is easier to destroy than to construct. Just as the history of art is that of creative artists, not of art critics, so the history of science and technology is that of original scientists and technologists, not of thinkers specializing in criticizing as severely as possible every tentative solution to a problem. The main moving force of the advancement of knowledge is not contradiction or criticism but inventiveness in facing and solving new problems. Criticism is the control of the process of inquiry, not its energy source. Like every other control it must be handled with care, for whereas in some cases criticism uproots error, in others it just gives vent to neophobia: witness the editorial process in learned journals. (Cf. Garcia, 1981.)

Criticism is necessary in all conscious endeavors, whether cognitive or practical, but not all criticisms are equally valuable. The best criticism is constructive: one engages in it not to show off (as is the case in carping) or

for sheer *Schadenfreude*, but for the love of truth or usefulness. It is part of the self-correcting mechanism of inquiry, be it in science, technology, or the humanities. (Ideology and pseudoscience lack such a mechanism. If only for this reason it is mistaken and dangerous to hold—as Feyerabend (1975) has done—that science is no better than magic or religion.)

To conclude this section. Criticism is neither the mark of rationality (Popper) nor the life of science (Poincaré), but a component of all inquiry. Therefore, to disregard or stifle criticism is to give free rein to dogma and charlatanry—whereas to exaggerate its importance is to forget that it is a means not an end. The clash of ideas is important to clarify and evaluate them, but not more important than the ideas themselves. And there is no such thing as the critical (or dialectical) *method*, i.e. a set of rules for sparking off clashes of ideas, examining them critically, or evaluating them, let alone for generating new ideas superseding the old ones. There are only a few general principles of argumentation, such as "Stick to the subject", "Do not tolerate self-contradiction", and "Concede defeat when it comes". But we are already trespassing on the next section.

3.2. *Controversy*

Controversy is criticism rendered public, at least within some inquiring community. Publicity has the great advantage that many can join in to contribute to the clarification of issues and the resolution of conflicts. But it also has the disadvantage of inflaming noncognitive passions and thus makes admission of error more difficult.

The history of religion, political ideology, and philosophy, is often presented as an unproductive sequence of indecisive fights over firmly held beliefs. In contrast, that of science and technology is usually presented as a progressive sequence of discoveries and inventions—all of them "positive" contributions—wherein controversy plays no important role. (See Agassi, 1963.) Moreover we are often told that controversy is a mark of vanity or, worse, an indicator of the immaturity of an epistemic field. Thus Kuhn (1970, pp. 6–7): "In a sense [. . .] it is precisely the abandonment of critical discourse that marks the transition to science".

There is much truth in the above. In fact controversy is less frequent, and far less indecisive, in science and technology than elsewhere; and it is less conspicuous in the older branches of science and technology than in the newer ones. (Thus there are far more and more acrimonious controversies

in psychology or sociology than in mathematics or chemistry.) Moreover, a sort of preliminary dogmatism is needed in science and technology, where one always starts by taking a certain body of knowledge for granted, never questions all of it, and replaces some of its components only if forced to.

Kuhn (1963) has pointed out the positive function of "dogma" in scientific research, holding that, as a consequence, resistance to change is normal and on the whole healthy. However, since the inquirer regards his background knowledge as only a launching pad that may have to be modified, it is incorrect to call it a dogma. His may be called *methodical dogmatism*, the companion of the *methodical skepticism* discussed earlier and the enemy of systematic dogmatism. Indeed the "preconceptions" or even "prejudices" that a scientist or technologist inherits are always on probation and may be modified or discarded altogether, some for having been shown to be erroneous, others for having lost value or interest. All scientific and technological (and also humanistic) knowledge has a built-in obsolescence. On the other hand ideological (in particular religious) dogmas are untouchable, and so are the principles of school philosophy. Those who reject them are branded as heretics and perhaps also punished. On the other hand those who criticize accepted scientific or technological principles or practices may be regarded as unorthodox but, if proved right, are hailed as revolutionaries or benefactors.

Just as excess controversy is a sign of immaturity of an epistemic field, so too little controversy may indicate senility, for every radically new idea is bound to elicit controversy if published. The healthy state of an epistemic field is one where most research is done along traditional (but fertile) lines, some explores radically new (but promising) approaches, and another argues the merits and demerits of either. No special control mechanism is needed to produce or keep such division of labor: most inquirers are naturally inclined to follow a successful tradition, only a few are capable of staging revolutions, and the controversy lovers, so numerous in the humanities, are scarce in science and technology. In fact most scientists shy away from controversy either from timidity or because they believe the positivist legend that only "positive" findings matter. As for technologists, they are even more circumspect though for different reasons: the design of a new artifact may be an industrial or military secret or, once in production, controversy over it could hamper its marketing or use. Not many technologists are as public spirited and courageous as Ralph Nader.

There is controversy in nearly every contemporary research field. Sometimes controversy indicates the inadequacy of, hence dissatisfaction

with, some received ideas or practices; at other times it indicates resistance to revolutionary innovations; and at still other times, clash with vested cultural, political, or economic interests. Whatever else it may indicate, controversy shows the interest aroused by the issue; in turn interest is a (weak) indicator of importance. (Nobody except Byzantine philosophers or theologians gets excited over trivia.)

Here is a quick sample of recent controversies in science and technology. (*a*) The controversy over the physical interpretation of the mathematical formalism of the quantum theories has continued unabated since the inception of these theories. The main rivals are the positivist (or Copenhagen) and the realist (or objectivist) interpretation. (*b*) The controversy between the "big bang" hypothesis of the origin of the universe and the steady state theory died out in the mid 1960s. However, a few astronomers do not accept either position and are looking for alternative explanations of the redshift of distant galaxies. (*c*) Although most biologists admit the fact of evolution, not all of them believe that it has been gradual or that it has proceeded exclusively on random mutation and natural selection. (*d*) Depending on which traits are taken to be peculiar to *Homo sapiens*, the origin of modern man is assumed to date back either 50,000 or 300,000 years ago. And whereas some investigators postulate a single line, others conjecture that there have been at least two (one of them the Neanderthal man). (*e*) Psychology is torn by the nature–nurture, the behaviorism–cognitivism, and the mind–body controversies. (*f*) In the social sciences the old battles between individualism and holism, as well as between idealism and materialism, are still reging—and these are only two out of dozens of spirited controversies. (*g*) Technologists are divided over the hardness of technology (advanced, appropriate, or soft) that should be used, over the accountability of technologists to society, etc. (*h*) Military experts debate endlessly over the merits and demerits of intercontinental missiles, neutron bombs, and other weapons.

None of these controversies is likely to be barren provided they are conducted with some rationality and in freedom. As it is, controversialists cannot help employ persuasive means—in particular rhetorical tricks and showmanship. As for freedom of inquiry, which includes the freedom to criticize and publish criticism, as well as to hold unorthodox views, it varies across research fields and from society to society, but it is never complete. Some groups, even within the scientific and humanistic communities, wield more power than others, and so are better equipped to impose their own views. In science most of the time there are illustrated despotisms, but

occasionally they stifle legitimate debate. For example, physicists are nowadays less tolerant to unorthodox views than social scientists.

Scientific and technological controversies eventually come to an end if they are purely technical, i.e. if they do not involve any philosophical or ideological principles. They come to an end in any of several ways. Here are some: (*a*) one of the parties exhibits a decisive observation, calculation, or technique; (*b*) a synthesis of the two rival views is proposed, that incorporates the good points of each; (*c*) a third party enters the controversy and wins the day; (*d*) the issue is forgotten for lack of interest; (*e*) censorship stops the discussion and imposes the official view: (*f*) one of the parties brandishes a bunch of computer printouts and announces 'The computer says so'.

There should be no doubt as to the efficiency of criticism to weed out error and thus clear the way for new ideas and even novel approaches. Thus Galilei's polemical works were no less important, for the revolution which he spurred, than his scientific ones. However, we should not follow Popper and overrate the power of criticism to advance knowledge, for the thrust of scientific and technological research is invention and discovery, not argument. In these fields controversy is the salt not the roast.

Besides, scientific or technological controversy, if overdone, can have negative effects. One is that it obfuscates the parties involved, which makes them lower their standards of rigor and objectivity, as well as harden whatever dogmatism may infect them. Controversialists are finally won by the desire to win the argument rather than to find the truth. Another negative effect of controversies is that they often contribute to the fragmentation of a research field. This happens whenever each party sees only one of the sides of a question, with its legitimate problem system, but fails to see the other. One example is the century-old debate, mostly mute, between the biologists who seek laws and those who try to establish life histories or phylogenetic lines. Far from being competing research lines, these are mutually complementary and should be merged instead of being kept separate.

The philosopher can participate in scientific or technological controversies either by taking sides or as an arbitrator—particularly if he feels that the contending parties are talking through one another and should cooperate rather than fight. And the historian's task is neither to hush up controversy for fear that it will spoil his story of cumulative induction, nor—*pace* Agassi's brilliant allegation (1963)—to read the history of science and technology in the same way as that of ideology or philosophy,

i.e. as a story of schools and their controversies. In science and technology controversy is healthy provided it is conducted honestly and in freedom, but it is seldom the main occupation of researchers. The main task of scientists is to unveil the world, that of technologists to design means to control it.

4. Concluding remarks

One way of gaining knowledge is by making observations or measurements, another is by forming and checking conjectures, and a third is by ferreting it out of available knowledge, i.e. by inference. All inference is from some nonempty set of premises—data, hypotheses, or definitions. True, the set of logical truths can be said to follow from nothing, i.e. to be entailed by the empty set of premises—provided rules of inference are granted. *Ex nihilo nihil fit.*

Inference yields new propositions but not new concepts. Thus if we know the sum and the difference of two numbers, we can find out what those numbers are—provided we have already the notions of number, addition, subtraction, and division. Therefore it is mistaken to call quarks, genes and mental processes 'inferred entities': they are hypothesized not inferred.

We need both rigorous (valid, deductive) and nonrigorous (invalid, nondeductive) inference in all fields of research. Even the most austere research, if fruitful, will beget logically illegitimate brain children—new concepts not defined in terms of the known ones or new propositions that do not follow from the known ones. No one type of inference is more productive than any other type. New concepts are not the product of inference, and hypotheses obtained by rigorous deduction (e.g. in theoretical physics) need not be truer or deeper than conjectures born of sheer insight.

Inference has been regarded as a surrogate of experiment (Rignano, 1923). To be sure some reasoning is thought-experiment and it may save us or, on the contrary, suggest real experiment. (Think of planning chess moves or of playing around with mental images or scale models of molecules, buildings or armies.) But no reasoning can replace experiment: there is no substitute for empirical data. And no experiment can replace reasoning: every experiment is designed and interpreted rationally. Reasoning and experiment are therefore mutually complementary, not interchangeable.

A hasty induction has led empiricist philosophers to proclaim the

inductivist thesis that every generalization, or at any rate every true (or warranted or believable) generalization is inductive: that we start every inquiry into matters of fact by studying individual cases, and conclude it by summarizing our findings into inductive syntheses, never proceeding from hypothesis to observation. By contrast rationalists, seduced by the success of pure mathematics, uphold the *deductivist* thesis that induction is unimportant or incidental: that we start every inquiry with some hypothesis or other, and find out by deduction what it commits us to; experience would come, if at all, at the very end, and its power would be negative rather than positive: it refutes but does not verify.

We endorse neither inductivism nor deductivism but do adopt elements of each. In our view research is sparked off by problems, not by observation or by hypothesis. Observation poses problems or checks hypotheses, but it is neither the only source of problems nor the only way of testing hypotheses. All inquiry except the most trivial involves deduction mingled with plausible inference, and the latter includes induction but is not restricted to it. Confirmation and refutation are seldom conclusive—except, again, in trivial matters. And induction supplies only relatively modest hypotheses and no theories at all. However, induction is involved in the checking of every generalization. The empirical worth of a hypothesis or theory is gauged not only by its resistance to refutation attempts but also by its positive confirmation or inductive weight (the set of data it accounts for) as well as by its theoretical support (the set of theories with which it jibes). But all this will be discussed at length in Ch. 12, after having studied the notions of problem, hypothesis, theory, explanation, forecast, and their kin.

PART III

EXPLORING AND THEORIZING

CHAPTER 7

EXPLORING

From an epistemological viewpoint, to live is to face and solve problems. We face practical problems, such as that of finding the next meal, and cognitive problems, such as that of finding out the nutritive value of a given food. We tackle some cognitive problems for pleasure and others because their solution may help us solve practical problems. Likewise we handle some practical problems for their own sake and others because their solution may allow us to pose or solve cognitive problems. All higher vertebrates handle cognitive problems, and man excels at inventing new problems on top of those he meets in daily life. And yet traditional epistemology has paid little if any attention to problems, the starting point and object of every inquiry. As a result we know very little about the biology, psychology, and logic of finding and solving problems.

In this chapter we shall peek at exploratory behavior and shall identify some problem sources. We shall also study the ways cognitive problems are selected for investigation. In the process we shall meet two concepts that in a way are mutually complementary: those of intuition and method. This will take us to the question of approaching a problem, which is part of every research. Finally we shall unveil the form and content of problems. None of this will transform the art of formulating problems into a science, but it may advance the art and help us understand the most fascinating of all cognitive operations, namely research.

1. Exploratory behavior

1.1. *Psychobiology of Problems*

Many animals seem endowed with an innate exploring or investigating ability (Pavlov, 1927, p. 12). This is the capacity to recognize and tackle problems of various sorts—e.g. "What is that?", "Is that edible?", "Is that friend or foe?", "Where may I find water?", "Can I negotiate that obstacle?", and the like. What is peculiar to man is not the capacity to recognize and tackle problems but that of formulating them consciously and explicitly, and investigating them methodically rather than by blind

trial and error. Also peculiar to man, at least over the past 10,000 years or so, is interest in purely conceptual problems removed from practical concerns, and pleasure in questioning solutions as well as in conceiving new problems.

A number of biologists seem to go farther and claim that all surviving organisms have "solved" successfully the "problems" posed by their environments, and that such "solutions" (which others call 'adaptations') constitute genuine inventions. Some even say that biological evolution is a problem solving process. And a few have gone as far as to hold that all a scientist or an engineer can do is to imitate, adapt or improve such "inventions". However, most biologists, if pressed, will adopt a sober attitude and own that all that is metaphor: that, although all organisms face challenges, not all challenges are or pose problems. And most psychologists will agree that only some animals, namely those capable of learning (and therefore knowing something), handle problems proper.

An animal may be said to *face a problem* when it feels the need or wish for some additional piece of knowledge—which of course presupposes that it did have some knowledge to begin with. The problem is *cognitive* if its solution might satisfy a purely cognitive need or wish. It is *practical* if its solution might be employed as a means for attaining a practical goal. In particular, a practical problem is *moral* if its solution might guide behavior that could affect the welfare of some conspecifics.

Like any other distinctions, these should not be transformed into detachments. Thus finding out that a certain drug has serious toxic side effects poses the cognitive problem of finding antidotes to it, or the moral problem of whether it should be banned. And appreciating the miseries of underdevelopment poses the cognitive problem of how best to contribute to alleviating them, as well as the practical problem of implementing whatever measures have been designed to solve those practical problems. In sum, sometimes sheer curiosity or wonder, and at other times need or want, sparks off inquiry, which in turn may result in discoveries or inventions capable of satisfying either. There is then no incompatibility between the culturalist and the economicist hypotheses concerning the root of inquiry: they are mutually complementary rather than exclusive.

An animal engaged in solving a problem explores: it explores the environment or itself, in particular its own store of knowledge and skills. However, not all exploratory behavior is problem solving: only *purposeful exploration* is. Thus the automatic and random exploration conducted by animals incapable of knowing anything is not of the problem solving kind:

not knowing anything, they have no knowledge gaps to fill. And it is doubtful that we are justified in assigning inquiring automata the ability to solve problems, for they have no purpose or design of their own: any intention they may have has been built into them by their manufacturers. They are not really inquiring machines but machine aids to human inquiry. As Duncker (1945) put it, "A problem arises when a living creature has a goal and does not know how this goal is to be reached".

Purposeful exploration, be it of the environment or of oneself, is triggered by the need or wish to cope with novelty, experience new situations, perceptions or feelings, or conceive new ideas. Neuropathology suggests that the arousal of curiosity and the drive to inquire are states or events in certain subsystems of the brain, in particular the mammillary bodies, the limbic system, and the hippocampus. In fact when any of these systems is damaged by accident or by drugs, in particular alcohol and marihuana, the questioning ability is seriously impaired. This impairment is extreme and irreversible in Korsakoff psychosis, characterized by lack of curiosity and initiative: the patient is apathetic, lacks initiative, and places himself at the mercy of his environment. He is good only at repetitive tasks: he cannot face new problems, let alone seek them.

A Korsakoff patient does not cope with novelty but at least he can be aware of it. In deep sleep or coma we may not even perceive, let alone seek, new stimuli: we are not alert. The alertness center of the mammalian central nervous system is the brainstem reticular formation (Moruzzi and Magoun, 1949). This system receives inputs from all sensors, and acts in turn on both the cerebral cortex and the endocrine system, in particular the pituitary-adrenal system. In turn, there is a feedback of the latter systems on the reticular formation, so that arousal is really a neuroendocrine process (Hennessy and Levine, 1979). Moreover the action of the reticular formation is nonspecific: it does not "say" 'Look at that' or 'Smell this', but just 'Watch out: something is up'. It prepares the higher centers of the brain to process the output of the sensors, or keeps them generating new information.

Exploratory behavior may be elicited by environmental events or by internal events, particularly in the brain. In either case exploratory behavior is driven by curiosity, which may be regarded as the drive generated by a knowledge deficit or imbalance. Thus a problem may be construed as the difference between what is known and what one wishes or needs to know.

Animals do not always wait for stimulation to trigger exploratory

behavior: all of the higher vertebrates actively seek stimuli of certain kinds, so that stimulation comes sometimes as a result of behavior, not the other way round. (They seek excitation up to a certain point, avoiding it only when it becomes excessive. Roughly speaking, the value of the overall external stimulation is bell shaped, but the maximum shifts with the internal state: to lower excitation values when the organism is tired or sick, to higher excitation values when it is healthy. (See Leuba, 1955.))

The schema of active exploratory behavior is not *stimulus–new internal state–response* but rather *curiosity–exploratory behavior–stimulus–new internal state-response* (Berlyne, 1954). Moreover, as exploration proceeds curiosity may be satisfied—i.e. the epistemic drive reduced. Or, on the contrary, the discovery or invention of new items may augment curiosity and thus sustain exploration or redirect it. Thus, whereas most other behavior types tend to reduce the corresponding drive, exploration may increase it. That is, inquiry is often self-sustained: it keeps generating problems. We call this type of inquiry *research*, to distinguish it from mere *fact-finding*, a process that reduces the epistemic drive as the corresponding problem is solved.

Looking for an address and computing a value of a function are instances of fact-finding inquiry. On the other hand conceiving of a model of the problem-solving process, and designing an experiment to test it, are examples of research. A fact-finding process comes to an end when a solution is found or shown to be impossible with the means at hand. In contrast, research generates new problems as it proceeds: in principle it is open-ended either because the solutions are imperfect or because they allow one to pose further problems. Most computer programs are of the fact finding type and a few of the research kind, but none is designed to pose new problems. Computers are problem solving aids, not problem finders.

A problem is a knowledge gap, and a problem solving process is one aiming at filling such gap: it has a very definite purpose. If we knew more about the neurophysiology of questioning and learning we would be able to picture this process with some precision, as the trajectory of a point in a cognitive space, from formulation to solution. See Figure 7.1. The starting point is included in the region of the cognitive space constituted by all the items relevant to the formulation of the problem—i.e. what may be called the *problem setting* or *framework*. Unfortunately, we do not even know which variables to take as axes for building such cognitive spaces.

Presumably the cognitive space of a newborn is very small, exclusively sensory-motor, and nearly amorphous. As the child grows up and learns,

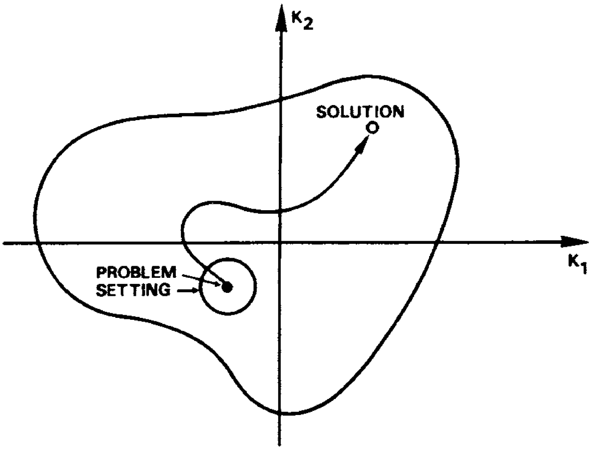


Fig. 7.1. Cognitive state space, and problem solving as trajectory of point representing state of knowledge, from formulation to solution. The axes K_1 and K_2 are (ranges of) functions representing neurophysiological variables characterizing the particular neural system performing the cognitive task in question. The cognitive states in the neighborhood of the problem constitute the problem setting, i.e. the collection of knowledge items relevant to the formulation of the problem.

its cognitive space expands qualitatively (new axes) as well as quantitatively (wider regions), and it becomes structured—i.e. the various knowledge items get interrelated forming systems. Such systems act as frameworks or settings for new problems. Thus when confronted with a new problem we are not always at a loss, but can often place the problem in our cognitive space and reconnoitre its neighborhood (setting) in search for clues. To change the metaphor, the inquirer forms, stores and organizes "frames" or data structures representing stereotyped situations, like moving about in a school building; such frames and frame systems are ready to receive or generate new experiences (Minsky, 1977).

All hard problems demand dedication, often over very long periods. However, it is common experience that continued exclusive concentration on a single problem does not pay: one is likely to fall into a rut that may lead to a *cul de sac*. Distractions are useful at this point: either turning temporarily to lighter problems, or just going for a walk. This manoeuvre has sometimes been interpreted as shoving the problem under consciousness and letting the Unconscious do the work. However, the recourse to the inscrutable Unconscious is not only a refuge in ignorance but also quite unnecessary, for there is a better explanation.

Concentration on a single problem tends to strengthen certain neural pathways while inhibiting others. (Recall Ch. 1, Section 1.1.) Distraction loosens some connections and gives others a chance to be established. This allows us to get off the trodden path and explore alternatives. When not under the pressure to move along the rut one may wander more freely in the cognitive space, which may eventually lead to the desired solution. Such happy end would not occur if we had no definite problem in hand (or rather brain) and some competence to handle (or rather mind) it. Neither the Unconscious nor blind chance (pervasive though it is), but just the leisurely exploration of new possibilities, does the trick.

In general, repeated experiences facilitate the formation of routines and *eo ipso* impair the ability to solve and even "see" problems of new kinds. Again, the use-disuse hypothesis explains this phenomenon: the neuronal connections become stronger, the more they are exercised. Hence experience, when confined to a few kinds, tends to decrease plasticity, whereas it tends to enhance it when varied. When a routine develops the tendency to disregard problems of new kinds, or to "assimilate" them to those of old kinds, it becomes an obstacle. One tries his old tricks on the new problems and, more often than not, one fails. This phenomenon of crystallization is what the Gestalt psychologists used to call *Einstellung* (problem solving *set*), and Duncker (1945) *functional fixedness*. (Cf. Katona, 1939; Birch and Rabinowitz, 1951.) People with fixed tasks become more and more rigid in their ways, until they become capable only of reproductive not productive thinking (Wertheimer, 1961). *Moral*: When scouting for someone capable of facing a tough new problem do not advertise for an experienced specialist but for a newcomer willing to take the risk and jump over disciplinary frontiers. He may undergo the neural reorganization we experience as the flash of insight (or *aha* experience).

By themselves, then, learning and praxis do not improve questioning ability or problem solving competence. The effect of both depends largely on the inquirer's attitude and opportunities. If his goal is to maximize his knowledge store or his practical utility, he will absorb solutions more readily than generate problems. On the other hand, if his goal is to maximize his contribution to knowledge, he will prefer unsolved problems, whether ambitious or modest. However, even the research-oriented inquirer—the one who hops from problem to problem instead of accumulating solutions—faces at all times the danger of becoming an erudite. Indeed, as science, technology and the humanities advance, it is only natural, as well as reasonable, to resort to the stock of solved problems

when first confronted with a new problem. This tactic is bound to succeed if the new problem is indeed similar to an older solved problem. But if the problem bears only a superficial analogy to an older solved problem, the erudite's procedure will lead nowhere or, worse, it will yield a wrong solution, or will generate more useless problems than it solves. (Adams (1980) calls this the 'hit-and-run' approach.)

The questioning ability, the capacity to "detect" new problems, can be trained or inhibited. The training can be of either the argumentative or the investigative type. The former consists in encouraging questioning and arguing without engaging in any inquiry beyond some fact finding—usually the spotting of the proper chapter and verse. This is the training that theology students, and even many humanities students, receive in the hope of discovering the truth by exegesis of canonical texts and by argument. As we have seen (Ch. 6, Section 3.1) it serves the entrenchment of dogma rather than the generation of new knowledge. The best training of the problem finding and solving capacity is one in mathematics, science, or technology. Indeed in these fields original work consists in investigating open problems with the aim of coming up with new results. Not that criticism is absent from such training, but it is only one among several means for producing or at least controlling new knowledge.

The questioning ability can be inhibited by repeatedly declining to answer the child's natural questioning, by giving unintelligible or false answers, or by imposing the uncritical acceptance of what is regarded as the right view. Unfortunately too many schools all over the world discourage questioning: they are *dressage* schools aiming at the training of conformists. (As cosmologist Hermann Bondi has stated, the few failures to this training in question-stopping are called 'scientists'.) There are three obvious reasons for this kind of schooling. One is that a questioning child is a source of additional work or even trouble. Another is the belief that all a child must learn are "facts" (data) and skills—questioning not being among the latter. A third is that questioning can be worse than annoying: it can be subversive, and most of the people who wield power do not want society to change (Ackoff, 1978, pp. 4–5). As a South American minidictator said, there is but one step from the scientific finding that matter is unstable, to the suspicion that even the most sacred institutions could be unstable as well.

Because of the obsession for results or findings, the problem of the psychological mechanism of the emergence of problems has hardly been studied. All we seem to know is that inquisitiveness is innate in many

species and that it has a positive survival value. On the other hand there are a few hypotheses about the problem solving process. The oldest is perhaps that problem solving proceeds by association. Although this view may explain (in a superficial way) why some ideas are recalled at the proper time, "it cannot explain the appearance of other ideas, e.g., original ideas" (Meier, 1931).

Most of the workers in this field are either empiricists or cognitivists. The former hold that, when tackling a problem, the animal proceeds by trial and error and gropes gradually towards the solution, i.e. reaches it through decreasing errors. This is at best a description of problem solving behavior. It does not explain the process because it does not tell us how the tentative solutions (hypotheses) are formed or why the process should converge. (A genuinely haphazard trial and error process should hit on the solution only by accident: there should be no gradual diminishing of discrepancies.) On the other hand cognitivism claims that the animal starts out its exploration with a whole set of hypotheses in mind, or else hits on the correct one in a flash of insight.

There may be something in each of the above rival views, and undoubtedly both of them miss something. The animal's behavior in the face of a new problem depends on its experience in handling problems of a similar kind, as well as on its goals and expectations. If it has no experience whatever, it will either avoid the problem altogether or tend to proceed by trial and error until reaching a satisfactory solution or giving up in fatigue or boredom. If on the other hand the animal does have some experience with similar problems, it will start out with a fund of guesses and try them out in succession. In either case, though, the subject keeps forming and testing new guesses. Moreover the process is not totally haphazard because there is feedback from some outcomes. I.e., unless the proposed solutions are totally irrelevant, or their adequacy kept from the subject, he may learn from his mistakes—not only learn to give up some of the wrong answers but also in what direction he should advance toward the correct response.

There is no finding and solving of problems without some motivation, but there are many kinds of motivation. Even in the case of purely cognitive problems, such as those in mathematics, basic science, and the humanities, the individual researcher may be driven not only by curiosity but also by love or hatred, celebrity or power, security or money. (My own motivation in writing my best physics paper was political.) So much so that, when any such noncognitive goal is reached, the individual may no longer feel motivated to pursue his research. (Most research careers terminate in a

doctorate, a professorship, or a deanship.) Yet no extracognitive drive, however strong, will accomplish anything of cognitive value unless accompanied by strong curiosity helped by talent, imagination, and industry.

Nor is a strong epistemic drive, allied to competence, sufficient to succeed in problem solving activities. Three further factors are involved: aiming at certain collective goals, choosing the right problem, and luck in investigating the problem at the right time in the right inquiring community. The collective goals are those of the given field of inquiry: they are not impersonal but are in the intersection of all the sets of individual motivations of the workers in the given field. Thus the collective goal of mathematics and basic science is the *finding of patterns*; that of technology, the *designing of useful artifacts*; that of the humanities, the *understanding of culture*; that of the liberal professions, *serving needs* of others. No matter how selfish a worker in any of these fields may be, he must aim at one of those collective goals if he is to make any contribution to it. So much for the collective goals of inquiry. As for choosing the right problem at the right time in the right circumstances, this matter deserves a separate section.

1.2. *Sources and Choices*

The traditionalist approach to learning presupposes that basic (or perhaps all of) human knowledge is locked in certain texts, whether sacred or secular. This is the approach adopted by religionists, schoolmen, hermeneutic philosophers, and doctrinaire ideologues. It follows that learning consists in studying and suitably "interpreting" such texts, and trying to solve every new problem by "applying" the suitable passages. (Failure in solving problems is blamed on the "interpreter" not on the texts.) In short, in this approach there can be no genuine new knowledge, and this because no new problems are admitted.

Modern workers in mathematics, basic science, applied science, technology, and the advanced sectors of the humanities, make no such presupposition and therefore proceed in a radically different manner. They presuppose, on the contrary, that nobody knows everything, that much of what little we know becomes quickly obsolescent, that our natural and social environments keep posing new challenges to be conceptualized as problems to be investigated, and that certain inquiries consist in posing and solving problems with no immediate practical roots or implications. In short, research-oriented inquirers know that *inquiry consists in investigat-*

ing problems. And they know that there are two problem sources: *external*, or originating in the natural or social environment, and *internal*, or originating in our own fund of knowledge—or rather in its gaps.

Talk of problem sources should be taken metaphorically not literally, for problems do not spring before us the way water springs from an artesian well. We do not discover problems the way we discover iron ore deposits: problems, like their solutions and, in general, all conceptual objects, are *made* by us. To be sure, not all environmental challenges arise spontaneously: some result from mismanagement of natural or social resources. However, man has lived for many millennia without transforming such challenges into those peculiar neural processes we call 'problems'. External challenges are occasions or opportunities, not causes of inquiry. Unless there are well equipped, trained and motivated brains, no challenge, however pressing, will elicit the conception of the corresponding problem, let alone of a research project to investigate it. Thus even now most people do not realize the gravity and urgency of the challenges to the very survival of the human species—starting with the arms race—so they keep postponing facing them, i.e. transforming them into problems to be investigated and solved.

So, it is simply not true that the environment, in particular society, exudes problems: individual brains, immersed in a suitable culture, create them. Nor is it true that every single problem we do manage to pose and solve is nothing but a response to challenges posed by "the historical development of productive forces and production relationships", as an early advocate of sociologism put it (Hessen, 1931, p. 203). Most problems in logic, pure mathematics, physics, astronomy, anthropology, history, and philosophy—to mention but a few fields of inquiry—are internal, not external. Workers in these fields handle mostly problems originating in what used to be called 'idle curiosity'—a sin to be shunned by every properly educated child. Let us praise and promote idle curiosity, for it prompts the most noble and, in the end, also the most useful of activities: inquiry. As Nobel laureate George Wald said, "The great questions are those an intelligent child asks and, getting no answers, stops asking. That is known as growing up." (Quoted by Teuber, 1966, p. 583).

Obviously, there is an enormous variety of internal problems as well as a number of kinds of stimuli and inhibitors to the problem finding activity. See Table 7.1.

Having emphasized that many problems in basic research originate in disinterested (idle) curiosity, let me hasten to acknowledge that others are

TABLE 7.1
Some stimulants and inhibitors of problem finding and solving

Stimulants	*Inhibitors*
Exploring new territory	Censorship (by schools or ideologies)
New facts to be checked or explained	Routine
Anomalies or contradictions	Complacency
Moderate criticism	Destructive criticism or indifference
Moderate competition	Rat race
Group cooperation	Infighting
Social rewards	Absence of social rewards

offshoots of practical problems, or have solutions that can be used to solve practical problems, whereas still others are elicited by environmental situations. An example of the first kind is the wealth of mathematical and scientific problems posed by technological developments; thus hydrodynamics owes more to aeronautics than the other way round. An example of the second kind is the utilization of biology in medical research and practice. An example of the third kind is social science: here the choice of problems is largely motivated by contemporary events. Let us take a look at this point.

Coleman (1980) has shown how the Chicago school of sociology, which flourished in the first quarter of this century around Robert Park, was mainly interested in the problems associated with the urbanization of North America (slums, ghettos, gangs, prostitution, etc.). Likewise the Columbia school built around Paul Lazarsfeld tackled some of the problems that arose in the 1930s, mainly the formation of a national market and of a national communication network. These problems called for a methodological shift, from the study of small groups (e.g. urban subcultures), to samples of large regions, sometimes national samples. Finally, the third or contemporary phase appeared with the beginning of social policies (as distinct from economic policies) in health, education, welfare, employment, etc., in the late 1960s. This research into social policies involved process evaluation, product evaluation, and even some large scale social experiments. It is definitely a shift towards applied social science. (Interestingly enough its leader, Professor Coleman, started out as a mathematical sociologist.)

So, there is two-way traffic between the internal and the external sources of problems. Besides, like every other human endeavor, problem find-

ing and solving are subject to chance. Sometimes chance events spark off problems or solutions, at other times they nip in the bud problem solving processes. A book on probability, seen by chance by a meteorologist, gives him the idea of investigating random processes in the atmosphere. A film on alcoholics gives a neurophysiologist the idea of pursuing a certain clue regarding taste formation and aversion. A chance encounter between a mathematician and a psychologist suggests to the former a new problem and to the latter a solution to his current problem. Computers are protected against such chance encounters: they are always programmed. So, they are neither led to truth nor misled into error in a haphazard way. Human brains, on the other hand, are biased cognitive roulettes.

The bias lies usually in the curiosity of the researcher or in the urgency of the problem. But sometimes fashion is involved; some problems are fashionable even when unimportant, whereas others are unfashionable even when important. One cause is obvious though unjustifiable: most middle class people like to be thought of as gentlemen or ladies of fashion. (Nowadays it is unfashionable to say that Bach and Wittgenstein are boring, that psychoanalysis and much of cosmology are unscientific, or that the abuse of computers is a threat to science.) Another cause is less obvious and thoroughly justified: science and technology must be up to date and, justifiably, this draws many people to work on recently posed problems.

Finally there is the question of timing. The alchemists posed too soon the question of the transmutation of metals, and the monetarists are too late in advocating a return to the competitive (free) market as the solution to the problems of the advanced capitalist economies. As Nobel laureate Sir John Kendrew said, "One has to ask the right question at the right time". But what is a right question? And can we say beforehand whether a question is good or bad, or must we wait and see what impact research on it has had? Let us deal with one question at a time.

Clearly, not all problems are worth the same. Thus finding out the neural mechanism of problem formation has a far greater cognitive importance than discovering a huge new deposit of mineral oil, but the practical importance of the latter may be far greater than that of the former. So, when ranking problems (or anything else) we should start by specifying the respect in which we are doing it: cognitive or practical. And we should keep in mind that such categorization may be only temporary, for some problems that start out being "academic" may turn out to be practically important (such as the problem of nuclear forces), whereas some practical

problems (such as whether or not trenches for warfare should be straight) may disappear before being solved.

Once this has been done, how shall we assess the relative importance of problems in the given respect? Will difficulty be a suitable measure, or perhaps the length of time that the problem has been around? The former is not a good measure, for difficulty is largely subjective: some good mathematicians experience considerable difficulty in performing elementary sums, whereas computers can perform calculations beyond the reach of any brain. Nor is length of time—particularly when measured by the number of pipes smoked by Sherlock Holmes—a good indicator of problem importance. Thus the four-color problem was solved only recently after two centuries of unsuccessful trials, but it is neither theoretically nor practically important: it is a stray problem and presumably its solution will not elicit a spate of other mathematical problems. We must look for alternative criteria.

We may start by classing problems into good and bad according to this definition: A problem is a *good* problem if (*a*) its solution would make some contribution to knowledge or human welfare and (*b*) it can be solved with existing or conceivable means. But of course not all good problems are important: some are trivial, others mediocre, and a few exceptional. We stipulate that a good problem is *important* if (*a*) it belongs to a problem system (rather than being stray), (*b*) it is deep, i.e. presuppose or is likely to upset basic principles or traditional behaviour patterns, and (*c*) its solution is likely to make it possible for a number of other people to pose further good problems.

By using the above definitions we can estimate a priori the value of problems. However, here as elsewhere our estimates may be erroneous. In any case three things are clear. One is that all research proposals are evaluated both a priori and a posteriori, and prior evaluation rests on some more or less tacit idea of problem worth. Second, if these ideas of problem worth were rendered explicit and discussed in the open, project evaluation would be more rational and objective, and therefore more fair. Third, very few researchers have a knack for "spotting" important problems, even fewer the ability and drive needed to solve them. Most researchers are better at problem solving than at problem having. (Worse, many researchers do not even seem to know that research consists in working on problems—so much so that, when writing their final reports, they seldom start by stating their problem.) Only exceptional investigators, such as Newton and Hilbert, are capable of drawing lists of important problems nourishing

several generations of workers. This requires flair as well as experience. (Cf. Bunge, 1962.)

Estimating the worth of problems with hindsight is easier than doing it a priori, but there are pitfalls here too. First, historians know that it is easy, perhaps unavoidable, to write "whig history"—i.e. to view the past against the background of our own problems and aspirations, and so to overrate certain accomplishments of our predecessors (mainly those which we have used) and underrate others. Second, even an objective indicator of worth, such as the volume of publications devoted to a problem, is not completely reliable, for there are always extrinsic factors, such as fashion, ideology, and availability of resources. Moreover in certain fields, such as philosophy, the relation between the importance of a problem and the volume of the publications devoted to it is inverse, not direct. Thus in recent philosophy trivial problems, such as the status of counterfactuals, the raven paradox, and the way to identify individuals across (logically) possible worlds, have attracted far more attention than the concepts of problem, partial truth, or attribute.

Although there are no rules for generating good problems, it is possible to give some helpful advice in this matter. First, *reexamine old problems*, for some of them may have changed just because their setting has altered, so that they can now be restated and solved in a different way. For example, the entire problem system of genetics was altered by the emergence of molecular biology and, in particular, the discovery that DNA is the genetic material. Also, practical problems "do not stay solved" (Ackoff, 1974) because society changes. Thus the automobile solved the urban transportation problem at the beginning of our century but is no longer practical.

Second, *examine critically* the known solutions to old problems, for there must be some flaws in some of them. (Enrique Gaviola, personal communication.) For example, by repeating a run of measurements or calculations, or reanalyzing a set of assumptions, one is bound to find something to correct or improve, generalize or discard. Thus we still do not have a satisfactory derivation of thermodynamics from statistical mechanics, or a good theory of liquids.

Third, *apply known methods or theories* to new problems, and check whether they still hold. If they do, you will have extended the domain of those methods or theories, and if they do not you may have generated a whole new problem system. For example, the Newtonian theory of gravitation, originally devised to explain the solar system, was successfully extended beyond it—until its own limits were discovered.

Fourth, *generalize* old problems: try introducing further variables or more complicated relationships among them. For example, introduce nonlinearities or inhomogeneities in equations, or impurities or external perturbations in concrete systems.

Fifth, *dig deeper*: try to obtain known results from assumptions concerning deeper levels. For example, try to explain psychological laws in neurophysiological terms, or historical trends in terms of social and economic laws.

Sixth, *look outside*: search for relationships to problems belonging to different fields. For example, explain chemical reactions as nonelastic scatterings of the reactants, or relate economic activities to political and cultural ones.

Seventh, *go out of fashion*: choose tough and interesting open problems that are not being currently investigated by others.

If all of these recipes fail, keep searching the literature, consult with others, ask someone for a problem, or take a holiday. Should all these moves fail after a couple of attempts, give up: this is far better than simulating research when actually doing nothing but keeping busy with books, instruments or computers.

An experienced and productive investigator needs no advice on how to generate new problems, for he stumbles on them all the time, and he may be in a position to give such advice. Problems seem to arise spontaneously in the course of his research, and to do so beyond his control and often to his own surprise. Sometimes the new problems are so numerous that he has to farm them out to assistants or students. For example, in measuring a magnitude a scientist may discover a fault in the experimental design or its implementation, which flaw generates the problem of overcoming it. In the course of the investigation of this problem he may hit on a new effect or a new technique, and so he may get sidetracked: the derivative problem may prove to be more pressing or even more interesting than the original one. Or in building a mathematical model of some thing or process, a scientist or technologist may formulate a new equation posing a new mathematical problem; conversely, in studying some mathematical work the idea may occur to him of trying the newly learned mathematics on one of his substantive problems that had proved to be intractable before. In sum, original research, unlike routine research, is an *open ended and self-sustained process*. If and when it does grind to a halt it is because people have lost their curiosity or the freedom to satisfy it.

2. From intuition to method

2.1. *Intuition*

When faced with a problem of a familiar kind we resort to our own fund of knowledge, or to the knowledge stored in other people's brains or in libraries. This will not suffice if the problem is of a new kind: in this case we need additional knowledge and some intuition or flair to guide us in the search for such knowledge. Intuition or insight is of course that proteic and ill defined ability to spot problems or errors, to "perceive" relations or similarities, to form concepts or hypotheses, to conceive of strategies or tactics, to design experiments or artifacts—in short to imagine, conceive, reason or act in novel ways. (Cf. Bunge, 1962.)

The novelty occurring in a flash of insight is of either of the following kinds: *radical* (unprecedented) or *combinatorial* (the result of combining previously known items). Radical novelty in ideation consists presumably in the formation of a psychon (plastic neuronal system) of a new kind, whether spontaneously or under the influence of the activity of some other psychons. On the other hand combinatorial novelty in ideation, i.e. association, consists presumably in the linking of two or more activities in so many neuronal assemblies (Hebb, 1949, p. 134). As long as the latter remain isolated from one another, the two ideas stay apart. But if, perhaps by chance, a bridge is formed between them, then the relation between the two formerly separate ideas emerges suddenly as a third idea. In both cases chance, as well as the deliberate scanning of a region of the cognitive space, are likely to occur.

The newborn has no insight: he must build it up. However, sheer quantity and variety of experience are not sufficient to acquire a powerful intellectual intuition although they are necessary for it (Birch, 1945). Hard work of the right kind, success, and luck—i.e. suitable circumstances—are needed as well. Hard work of the right kind—tough tasks with scant means—is necessary for it cannot be performed in a routine fashion but calls for imagination. Success is needed because repeated failure discourages and counsels turning to easier tasks. And luck is needed because adverse circumstances beyond our control may prevent us from honing our wits. Thus too long hours applying rules of some kind, or making routine calculations or measurements, will blunt one's intuition. On the other hand repeated exposure to varied problems requiring more ingenuity than specialized knowledge will hone it. For this reason Dreistadt (1969)

Fig. 7.2. The Farm Problem: divide the L shaped lot into 4 parts with the same shape and size.

recommends proposing problems such as the Farm Problem to cultivate problem solving skills. See Figure 7.2.

No problem can be understood and handled successfully without some intuition or "vision" of it as a whole and against its setting. This holds for all problems, whether cognitive or practical, in every field of human endeavor, from science to technology, from art to the humanities, from management to politics. However, the need for intuition in every field of human endeavor does not entail that is suffices, or that we cannot go beyond it, or that it is superior to experience and reason.

Intuition is insufficient: it is not a substitute for work but a guide to it. There is no substitute for hard work except of course soft work—but this one does not really work. The so-called "aha experience" (or "eureka phenomenon") does not come out of the blue but after hard thinking or doing, after trying different hypotheses or methods, and repeatedly reorganizing one's ideas. (Cf. Gruber, 1981.) Divination is no short cut either: it is no substitute for exploration in the light of specialized knowledge. For example, the rate of success of dowsers (water witches) is far inferior to that of geologists and water engineers: "We know of no other hypotheses that have been put forth so persistently over such a long span of years with such surprisingly negative experimental findings as the hypothesis that water witching 'works'" (Vogt and Hyman, 1959, p. 82; see also Randi, 1979).

Intuition is preliminary and tentative rather than final: it must be elaborated and checked. Moreover mathematics, science, technology and philosophy keep introducing counterintuitive or implausible concepts, theories, and methods. (Actually every radically new idea, even if initiated intuitively by its creator, is likely to sound counterintuitive to all others.) In short, intuition is indispensable but intuitionism, the epistemological doctrine according to which it is sufficient, is false. (For a criticism of philosophical intuitionism see Bunge, 1962.)

Counterintuitive ideas are particularly conspicuous in contemporary science and technology. This is so because appearance is only skin deep, and reality so complex that it can be grasped and controlled only with the help of very sophisticated instruments. Photons are not tiny particles, solid bodies are not continuous, organisms are not machines, and national economies are not households. Actually we need not go far out into modern science and technology to find problems and solutions that defy intuition. Recall the story of the king who promised to reward one of his subjects with a few grains of wheat: one for the first square of the chess board, two for the second, four for the third, eight for the fourth, and so on. He had not realized that the total number of grains promised was $2^{64}-1$, which is more than the annual world wheat crop.

As long as we stay on an intuitive level we are constrained to proceed by guessing, also known as the trial and error "method"—actually the very antithesis of a method. To be sure, this procedure is bound to succeed sooner or later in the case of problems involving a manageable number of alternatives. Thus Edison and his coworkers hit on the right material for manufacturing light bulb filaments only after having tried haphazardly nearly 10,000 different materials—practically all those known in his time. (He could have proceeded methodically if solid state physics had existed in his time.) But the hit-or-miss procedure is unlikely to succeed with complex problems even after repeated trials. Electrodynamics and the dynamo were not outcomes of blind trials but of hard and sustained theoretical and experimental work. The alchemists tried for thousands of years to turn lead and other "base" metals into gold by combining and treating at random a number of substances. Nothing of the sort could possibly have worked: a wholly different approach was needed to solve that problem. We shall elucidate the obscure and important notion of an approach in Section 3.2 after having examined the concept of a method.

2.2. *Method*

We are all familiar with a number of methods, or regular procedures for handling problems of restricted kinds, in all sectors or human endeavor, from mathematics to butchery. For example, neuroscientists use the Golgi staining technique, in conjunction with microscopy, to render individual neurons visible. They also employ microelectrodes to record the electrical activity of single neurons, and electroencephalography to record the activity of entire regions of the brain. All these are *techniques*, or special

methods. Every technique has a restricted scope, i.e. it cannot be exported to distant fields of research.

Philosophers have been acquained since Antiquity with three conceptual methods: Socrates' maieutic method, Porphiry's tree questioning, and the method of hypothesis. (We do not rate contemplation as a method, for there are no rules for contemplating in a productive way.) The maieutic method supposedly consists in extracting knowledge, by clever questioning, from the student's inborn storehouse of knowledge. Actually the teacher must know the answer and must know how to supply clues that will allow the student to exploit his own background knowledge. The method works for purely logical problems but does not help solve problems in, say, statistics, biochemistry, or history. Tree questioning consists in surveying the set of possibilities and dividing it step by step into mutually disjoint subsets until the wanted subset is reached. It works whenever the subject is initially given a finite set of possibilities and has the means to eliminate the wrong ones—which is not often the case in original scientific or technological research. Finally, the method of hypothesis, or hypothetico-deductive method, consists in the sequence *invention* (of idea or device)—*trial*—*evaluation*. This method is used in all cognitive fields, including ordinary knowledge. And it is used systematically in mathematics, science, modern technology, and philosophy, where hypothetico-deductive systems (theories) are plentiful. See Figure 7.3.

Not all techniques are alike in accuracy, foundation (or justification), and cost. For example, to detect defects in the hull of a ship, X-rays and ultrasound are more accurate and better founded than banging with a hammer—but they are also more expensive. Again, forecasting the state of the economy by divination or on the basis of some economic ideology is inferior to extrapolating recent trends; and forecasting on the strength of well-confirmed theories and hard data should be superior to trend extrapolation.

A technique or special method may be called *scientific* if (*a*) it is intersubjective in the sense that it gives roughly the same results for all competent users, (*b*) it can be checked or controlled by alternative methods, and (*c*) there are well-confirmed hypotheses or theories that help explain, at least in outline, how it works. A method that complies with only one or two of these conditions may be called *semiscientific*, and one that complies with neither *nonscientific*. Note that, whereas condition (*b*) is that of testability (checkability), clause (*c*) is that of foundation or justification, as opposed to faith or authority.

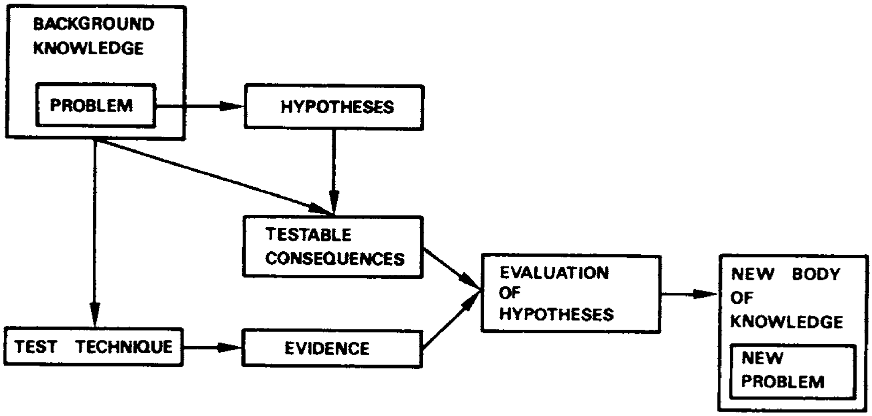


Fig. 7.3. A scientific research cycle. The importance of a scientific research is gauged by the changes it induces in our background knowledge or by the new problems it poses. Adapted from Bunge (1967a).

For example, whereas random sampling is scientific, the ink-blot test is not. Indeed the former is objective, can be checked by investigating the whole population, and is founded on the probability calculus. On the other hand the ink-blot test is subjective and mysterious. A good example of a semiscientific technique is the memorization of nonsense syllables, widely used to study verbal learning. It is objective and can be checked by alternative techniques, but it rests on a false hypothesis. The hypothesis is that, by memorizing nonsense syllables such as *giz* and *toc*, the "blank slate" condition is reproduced, so that learning does not depend on previous knowledge. But this is false: in order to memorize such syllables the subject is likely to associate them with familiar words or images—e.g. 'giz' with 'gizmo' and 'toc' with 'woodpecker'. There is no pure verbal learning. Like all learning, verbal learning is rooted in previous knowledge: it is a growth process, not one of haphazard piling up of unrelated items.

The inception of a new technique may revolutionize a field of human endeavor or it may even give rise to a new field. Think of triangulation, printing, the microscope, chromatography, or computers. The design of a new technique is, increasingly, the work of a sequence or of a team of specialists composed of scientists, technologists, and craftsmen. Much effort is being devoted in science and technology to the invention, improvement, and adaptation of new techniques. However, it should never be forgotten that techniques are means not ends: means for knowing things

or getting things done. The goal is either a bit of knowledge or a commodity.

The learning of techniques is an important part of the training of experts. It is also the easiest part, for one need not know why or even how the technique works in order to use it competently. (For example, nobody seems to know how the Golgi staining technique works.) Therefore most specialists are skilled technicians: performers rather than musicians, color mixers and dabblers rather than painters, microscopists rather than cellular biologists, electrode pushers and recorders rather than neuroscientists, calculators rather than mathematicians, and so on.

When on their own, technicians are likely to engage in routine work, haphazard tinkering, or even wasteful activity. Technicians perform best when they are components of teams led by productive artists, technologists, scientists, or managers. This is so because in such case their work services a long-range goal and receives the feedback necessary for correcting false moves. As a component of such a team a technician's tinkering may lead on to a new technique or to the discovery of a previously unknown category of facts.

Some techniques born with a narrow scope have evolved eventually into general methods. Microscopy, the method of successive approximations by iteration, and even the scientific method are cases in point. Indeed microscopy is now being used not only in biology, where it was first applied, but also in physics and various branches of engineering (e.g. metallurgy), as well as in some branches of the humanities (e.g. art history). The method of successive approximations by iteration is employed in all of applied mathematics and is a paragon of scientific procedure. One starts by computing or guessing a first approximation. This approximate solution is then fed back into the original problem to yield a second solution, and so forth. In this way more and more accurate solutions are obtained: the solution is the limit of the sequence. Every partial solution is built upon the preceding ones, and the increase in accuracy or degree of truth is uniform. That is, the iterative procedure is constructive and convergent.

We mentioned the scientific method as one with a large scope, in contrast with any of the special methods or techniques. What is it? Opinion varies from radical skepticism, which denies its existence, to dogmatism, which believes it to be a set of simple, invariable and infallible recipes for finding definitive truths. We take the *scientific method* to be the following ordered sequence of cognitive operations:

1. *Identify a problem* (whether gap or indentation) in some body of

knowledge—if possible an important bit of ignorance. If the problem is not stated clearly, go to the next step, otherwise to the subsequent one.

2. *State the problem clearly*, if possible in mathematical terms or in terms of measurement operations.

3. *Search for information, methods or instruments* likely to be relevant to the problem—e.g. empirical data, theories, methods of measurement or computation, measuring instruments, etc. I.e. scan what is known to see whether it can help solve the problem.

4. *Try to solve the problem with the help of the means collected in the previous step*. Should this attempt fail, go to the next step; if not, to the subsequent one.

5. *Invent new ideas* (hypotheses, theories, or techniques), *produce new empirical data*, or *design new artifacts* that promise to solve the problem.

6. *Obtain a solution* (exact or approximate) of the problem with the help of the available conceptual or material means.

7. *Derive the consequences* of the solution thus obtained. If the solution candidate is a hypothesis or a theory, compute predictions; if new data, examine the effect they may have on existing ideas; if new artifacts, assess their possible uses and misuses.

8. *Check the solution*. If the solution candidate is a hypothesis or a theory, see how its predictions fare; if new data, try to replicate them using alternative means; if new techniques or new artifacts, see how they work in practice. If the outcome is unsatisfactory, go to the next step, otherwise to the subsequent one.

9. *Correct* the defective solution by going over the entire procedure or using alternative assumptions or methods.

10. *Examine the impact* of the solution on the body of antecedent knowledge and state some of the new problems it gives rise to.

Table 7.2 exhibits schematically the scientific treatment of three typical problem types: one theoretical, one experimental, and a third one technological.

The scientific method has some points in common with the method of successive approximations by iteration. First, scientific research proceeds *gradually*: even the correct insights that may be hit upon by chance, and revolutionary new ideas, are the outcome of previous work, as well as subject to correction. Second, scientific research yields, at least with reference to the world of facts, *partial truths* rather than complete and accordingly final truths. Third, the scientific method, in contrast to the hit-or-miss procedure of common sense and unbridled speculation, is *self-*

TABLE 7.2
Three typical problem types.

Step	*Empirical problem: measure*	*Theoretical problem: explain*	*Technological problem: design*
1	What is the value of X?	Why does X have the value x?	How can X be done?
2	What is the measured value of X to within error e?	What premises entail that X is worth x?	How can we design a Y capable of doing X with maximal efficiency?
3	What experimental arrangement(s) Y allow to measure X with error less than e?	What theories Y, subsidiary hypotheses h, and data d imply that the value of X is x?	What information can help design Y?
4	Perform measurement of X with means Y. If Y proves inadequate go over to step 5, otherwise to 6.	Compute the value of X with the help of Y, h and d. If the result is implausible go to step 5, otherwise to 6.	Design Y with the help of the information gathered in previous step. Should design prove inadequate go to 6.
5	Try new technique Y'.	Try new theory Y', or new subsidiary hypotheses h', or new data d'.	Try new design.
6	Use Y' to measure X.	Compute the value of X with Y', etc.	Build an artifact (or a scale model) Y' or run a computer simulation based on the design.
7	What does the outcome of 6 imply or suggest?		
8	Measure X with alternative techniques. If the new result is implausible go to step 9, otherwise to 10.	Compute X with alternative assumptions or methods. If the new result is unsatisfactory go to step 9, otherwise to 10.	Put Y' to some uses and misuses. If Y' does not work satisfactorily go to step 9, otherwise to step 10.
9	Look for systematic errors and correct them.	Look for possible sources of error and correct them.	Look for flaws in design and correct them.
10	How does the new result affect knowledge or practice, and what new problems if any does it pose?		

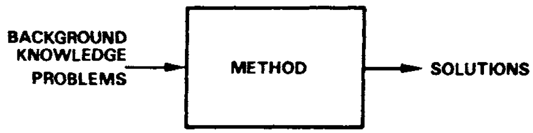


Fig. 7.4. A method as an input-output box. Input: problems and background knowledge; output: solutions.

correcting: it can help recognize errors and obtain higher order approximations, i.e. truer answers. One difference between the scientific method and the mathematical iterative procedures is that the former never allows one to find exact solutions of nontrivial problems. Another difference is that the scientific method can be applied to all inquiries, whether mathematical or empirical, scientific, technological, or humanistic. (Some philosophers deny this: we shall come back to this point in Section 3.2 and in Vol. 6.)

Every method, regardless of its scope, can be construed as a device transforming problems, together with the data relevant to them, into solutions. See Figure 7.4. In this context we understand by 'data' not only the givens of the problem but also all the components of the body of relevant available knowledge, including ontological and epistemological principles. And by 'solution' we understand the set of solutions of the given problem; thus the two roots of a second degree equation constitute the solution to the latter.

(A method M may be construed as a partial function from the cartesian product $2^B \times \Pi$, where B is a portion of background knowledge and Π a set of problems, into the set S of solutions. The solution to $\pi \in \Pi$ is $s = M(b, \pi)$, where $b \in 2^B$. We take the power set 2^B of B, rather than B, because every method calls for a whole set of bits of knowledge. The *M-soluble problems*, or problems soluble with the help of M, constitute a subset Π_0 of Π, such that the restriction $M_0 = M \mid \Pi_0$ is a bijection—i.e. one problem—one solution and conversely. In principle there is at least one method for every problem, and different methods may yield different solutions. Two methods are *equivalent* if they yield the same results. But of course they do so by introducing different intermediate steps between problem and solution. In the following diagram two alternative intermediate steps are shown.

In this commutative diagram $M_1 = F_1 \circ G_1$, $M_2 = F_2 \circ G_2$, and $M_1(\pi) =$

$$\begin{tikzcd}
 & I_1 \arrow[dr, "F_1"] & \\
2^B \arrow[ur, "G_1"] \arrow[rr, "M_1", "M_2"'] \arrow[dr, "G_2"'] & & S \\
 & I_2 \arrow[ur, "F_2"'] &
\end{tikzcd}$$

$F_1(G_1(\pi)) = F_2(G_2(\pi)) = M_2(\pi)$. A *tactics* is thus an ordered pair $\langle F, G \rangle$ such that $F \circ G = M$.)

Knowledge of methods is called *methodological*, and its complement *substantive*, knowledge. These two are distinguishable but not detachable, for the former is only a means for the latter, and it cannot be improved except with the help of the latter. (Imagine trying to design a new particle detector without knowing anything about particles.) Yet from the 17th century on there have been enthusiasts and even cultists of method, claiming that there is a method for every activity, and if only one hits on the right method he will perform to perfection and without any need for substantive knowledge. (An early and delightful exposition of this creed, methodolatry, was Bacon's *New Atlantis.*)

Methodolatry was badly shaken by the realization that (*a*) we know how to do a number of things without being able to formulate any explicit rules for doing them, and (*b*) we know a number of methods that often fail. Loving and building theories exemplify the former, and weighing and making statistical inferences are instances of the latter. (Recall the distinction between explicit and tacit knowledge: Ch. 2, Section 2.2.)

Methodoclasm, the opposite of methodolatry, holds either that there are no methods at all or that they are all equally fallible. This belief is even more wrong than methodolatry, for (*a*) actually there are a number of methods that work reasonably well, and (*b*) all methods, or at least the ways (tactics) of applying them, can be improved. Hence whereas the study of methodology does not produce instant scientists or technologies, its ignorance does not help discover and correct errors.

The advice often given by methodoclasts, namely "Study the history of knowledge instead of methodology", is only a partial answer, particularly because every historian writes from his own perspective, which may be more or less perceptive and tied to a more or less adequate methodology.

We do need methodology in addition to descriptive epistemology and the history of science and technology, if only because (*a*) people do not always think correctly—e.g. most of us tend to disregard unfavorable evidence and write off criticism, (*b*) there are thousands of effective techniques for analyzing statements and arguments, as well as for producing and checking data—not to speak of the uncounted methods for handling problems in pure and applied mathematics, and (*c*) the scientific method embodies certain rules of conduct that constitute the core of the ethics of science, and make it possible to search for truth.

Finally, a word on the value of methods. Each method is valuable to some extent (possibly nil) for solving problems of a given kind and with a certain aim. Change problem or goal and the method may cease to work or it may work better. (In other words, the valuation function for methods is defined not only on the set of methods but also on the sets of problems and goals.) Consequently when comparing methods we must take all three aspects into account. Thus it may happen that one method for solving a given problem is better than a second method with regard to accuracy but worse with regard to speed; or that one method is more general but also less simple than another. For example, the scientific method is the most general and accurate of all methods of inquiry, but it is also the one demanding the largest investment in knowledge and equipment.

3. Approach and Research

3.1. *Approach*

An approach is—roughly and metaphorically—a way of looking at things or at ideas, and therefore also of handling problems concerning things or ideas. Take for instance the matter of pest control. The chemical approach to it results in spraying pesticides regardless of any undesirable long term consequences: it is determined by a narrow viewpoint and short term goals (saving this year's crop or favoring the chemical company). On the other hand the biological approach to pest control is determined by a broad view as well as long term goals, and it results in trying multiple biological means, such as protecting or spreading certain species-specific predators (e.g. Lady Bugs), and developing pest-resistant strains.

We shall distinguish seven broad approaches to the study and handling of things or ideas: the vulgar, doctrinaire, mathematical, scientific, applied, technological, and humanistic. The *vulgar approach* rests on ordinary

knowledge, tackles both cognitive and practical problems, employs daily life procedures (in particular authority and trial and error), and is mainly interested in practical results. The *doctrinaire approach* rests on some ideology, tackles both cognitive and practical problems, resorts to authority, criticism and argument, and is mainly interested in practical results (including personal salvation and the preservation of the ideology).

On the other hand the *mathematical approach* is characterized by a formal basis (logic and mathematics), formal problems, conceptual methods, and the finding of patterns as well as the erection of formal theories. The approach of *basic science* rests on a fund of mathematical and scientific knowledge, as well as on the scientific world view, it deals with cognitive problems, employs scientific methods (in particular *the* scientific method), and aims ultimately at understanding and forecasting facts with the help of laws. The approach of *applied science* shares the basis and methods of basic science, but tackles only special cognitive problems, and aims at supplying the cognitive basis of technology. The *technological approach* is like that of applied science, but its basis includes also the fund of technological knowledge, and its aim is the control of natural systems as well as the design of artificial ones. Finally, the *humanistic* approach is based on the body of knowledge concerning human culture, it handles problems concerning intellectual and artistic culture, uses conceptual methods, and aims at understanding its referents.

In general, an approach may be construed as a body B of *background knowledge* together with a set P of problems (*problematics*), a set A of *aims*, and a set M of methods (*methodics*):

$$\mathscr{A} = \langle B, P, A, M \rangle.$$

Needless to say, every component of this quadruple is to be taken at a given time, and so is a (time dependent) collection rather than a fixed set. (Note that an approach is a projection of a conceptual framework as defined in Ch. 2, Section 4.1.)

If the background knowledge is restricted to the general philosophical framework or world view, we obtain three main types of approach: the atomistic (or individualistic or analytic), the holistic (or synthetic), and the systemic (or analytico-synthetic). The *atomistic approach* rests on an atomistic ontology, according to which the world is an aggregate of units of a few kinds, and a reductionistic epistemology, according to which the knowledge of the composition of a whole is both necessary and sufficient to know the whole. The goals of atomism are the same as those of science, and

the atomistic methodics boils down to analysis into components (or the top-down method). The *holistic approach* rests on a holistic or organismic ontology, according to which the world is an organic whole that may be decomposed into large partial wholes that are not further decomposable, and an intuitionist epistemology according to which such wholes must be accepted and grasped as such rather than analyzed and tampered with. The goal of holism is to emphasize and conserve wholeness and emergence, and its method is intuition. (Actually intuition is nonmethodical: recall Section 2.1.) Finally, the *systemic approach* rests on a systemic (or systems-theoretic) ontology according to which the world is a system composed of subsystems, and an epistemology that combines realism with rationalism. Its objectives, like those of science and technology, are to understand, predict, and control. And its methodics includes analysis as well as synthesis, generalizing and systematizing, and empirical testing. (For details see Bunge, 1977a, 1977b, 1977f).

Each approach can handle only certain problems: every approach is partly characterized by its own problematics. Thus the atomistic approach can handle only questions regarding individual behavior: since it does not recognize wholes with emergent properties, it sees no point in looking for patterns of global behavior, i.e. for laws of systems as wholes or units. Likewise the problem system of holism is limited: it is hardly interested in finding out, say, the atomic composition of a molecule, or the mechanisms whereby a group of people form or dissolve a society. By contrast, systemism, or the systems approach, keeps the positive aspects of atomism and holism, and it is the approach compatible with the scientific approach. Actually it differs from the latter only in that it presupposes far less substantive and methodological knowledge.

The advantages of the systemic approach are best appreciated by comparison with its rivals. Whereas systemism treats every system as a whole composed of individuals held together by bonds (part of the system structure) and immersed in an environment, atomism neglects both the environment and the structure; holism neglects the composition of both system and environment, and does not analyze the structure; environmentalism focuses on the environment at the expense of both composition and structure; and structuralism neglects both the composition and the environment, not to speak of the history of the system. Only systemism does justice to all three aspects and their changes.

Failure to adopt the systemic approach to the study or design of systems is bound to result in failure to solve the problems concerned or, worse, in

creating unnecessary new problems. For example, a marine biologist studying the migration patterns of sea lions will not make much headway unless he also studies the migration patterns of their prey. An evolutionary biologist studying the foraging abilities of honeybees will do well to study at the same time the evolution of the flowering plants preferred by bees. The archaeologist and the historian of technology cannot restrict their attention to artifacts, but must try and embed them (conjecturally) in the society that used them. The epistemologist is bound to be superficial or even utterly mistaken unless he keeps in mind that, although cognition is a brain process, it continues a tradition and it is now stimulated, now inhibited by social relations. The engineer will not succeed in designing useful large scale projects, such as dams and transportation systems, unless he cooperates with social scientists. The statesman will not introduce useful new legislation unless he realizes that every subsystem of society (the economy, the culture, and the polity) has strong links with the other subsystems of it. Although all this sounds obvious, some of it has been discovered only in recent years, and some is still to be learned by many specialists. Indeed, one may speak of the systemic *revolution*, or of the systemic approach as a new *paragon* (or paradigm). However, for the most part it is a silent revolution with but a few outspoken paladins.

In many cases, particularly in practical matters, the systemic approach violates some of the artificial frontiers between the various fields of inquiry. For example, the so-called new archaeology, that employs the systemic approach, merges history, sociology, and anthropology (cf. Sherratt (Ed.), 1980). And the new technology, that takes into account the social dimension of technical systems, is closely associated with the social sciences (see the journal *Technology in Society*). In all such cases unidisciplinarity is given up in favor of *interdisciplinarity*. In the latter approach the variables typical of a number of disciplines are not only taken into account but also interrelated: the resulting model is neither a single picture nor a mosaic of pictures but, to continue the metaphor, a sculpture. By contrast, the *multidisciplinary* approach consists in the simultaneous application of a number of unidisciplinary approaches: the result is a mosaic of pictures or models. (For an alternative analysis of these and other concepts see Jantsch (1972).) In any event, the problems concerning multilevel systems—and all problems concerning humans are of this kind—are not inherently unidisciplinary, multidisciplinary, or interdisciplinary: only our approaches to such problems can be uni, multi, or interdisciplinary. See Figure 7.5. We shall return to this theme in Ch. 14, Section 3.

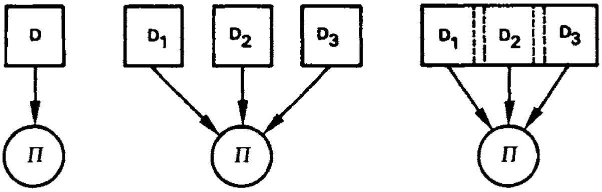


Fig. 7.5. Uni, multi and interdisciplinary approaches to a set Π of problems.

3.2. *Research*

We may speculate that understanding a problem is identical with the activity of certain psychons which, in turn, activate further psychons likely to do the job (Bindra, 1976). Thus if asked to multiply 13 by 67 the subject may choose to do so mentally, with paper and pencil, or with a calculator. In each case different psychons are likely to be activated: which ones depends on the subject's talent and experience, as well as on the instructions he has received. Likewise with problems that require research. Thus if asked to find the law of a certain process, the subject may—depending on his knowledge, intelligence, habits, opportunities, etc.—initiate a bibliographic search, or design an experiment, or sit down and speculate hoping to guess the law on the strength of a few clues such as data and analogies. In each case different psychons are likely to be activated sequentially. As the research proceeds some entirely new concatenations may occur. And, whereas certain psychons will interfere constructively with those in charge of the process, others will interfere destructively, resulting in sidetracking or halting.

Work on a routine problem may be assumed to consist in the sequential activation of a fixed set of psychons. (Analogy: groove or orbit.) Thus when psychon 1 formulates the problem, it activates (with probability near 1) psychon 2, and so on, until the target psychon—the solution—is reached. On the other hand, if the problem is new to the subject, then instead of a linear sequence or chain, a zig-zag process occurs, in which the connections between the various psychons are probabilistic rather than causal, and where the target psychon may never be activated, either because of the subject's personal failure or because the problem has no solution.

Schematically, the routine problem solving process looks like this: *Formulation–search–solution–checking*. The trademarks of this process are

definiteness of formulation and near certainty of solution. On the other hand the research problem solving process looks like this: *Raw problem–choice of approach–formulation of problem within this approach–search–research plan–solution candidate–checking–evaluation.* The trademarks of this process are indefiniteness of formulation and uncertainty of outcome. (Nobel laureate Peter Kapitza (1979): "The main attraction of scientific work is that it leads to problems, the solutions of which cannot be foreseen.") In most cases the outcome proves unsatisfactory, and the process has to be repeated either partially or *da capo*, from the very choice of approach or at least from the formulation of the problem within the given approach.

There are no cut-and-dried rules, e.g. algorithms, for performing most of the operations involved in a research problem solving process. At most there are heuristic maxims and methodological rules that help design research plans. Original research, i.e. work on research problems, is zigzagging, not straight; it is chancy, not certain; and it calls for direct cooperation (e.g. of technicians or consultants) or indirect cooperation (e.g. bibliography or referees). Therefore the idea of designing a computer program embodying a "general problem solver" (Newell and Simon, 1972) is wrong-headed.

The most one can do to expedite research is (*a*) to remove obstacles (administrative, ideological, etc.); (*b*) to supply human and material resources, and (*c*) to bring to light some strategies and tactics employed in successful research, and try and clarify, justify, and systematize them—i.e. to cultivate methodology. The strategies of inquiry are the general research methods, such as the scientific method and the methods of random sampling and of mathematical modeling. As for the inquiry tactics, they are of two kinds: techniques (special methods) and heuristics. Whereas techniques are reliable procedures, heuristics are suggestions that may or may not work—such as looking for analogies with problems of known solutions, or trying to generalize special results.

The most general strategy for handling research problems, be it in science, technology, or the humanities, was indicated a while ago, but it bears repetition. It consists in the following sequence of operations:

1. *Spotting the problem(s).*
2. *Choice of approach* (which includes background knowledge, methodics, and goal).
3. *Formulation of problem* within chosen approach (i.e. transformation of raw initial problem into well defined problem).

4. *Search for existing relevant information* (i.e. scanning of background knowledge stored in theories, tables, data banks, etc.).

5. *Design of a research plan.*

6. *Implementation of plan* (conceiving hypotheses, building theories, making computations, performing measurements, designing techniques or artifacts, etc.).

7. *Trying* (checking) the outcome of previous step.

8. *Evaluation* of candidate solution (idea, data, artifact, or what have you) in the light of the background knowledge as well as the outcome of the previous step.

9. If solution is found acceptable, *extend or revise* background knowledge.

10. Otherwise *start research process again*, either replicating it or modifying some components (e.g. hypotheses or methods).

Note again the difference between a routine and an original research project. The former ends up by answering questions, the latter poses further questions even when it attains answers. (In other words, whereas routine problems are information sinks, original problems are seminal. Yet not all sources are sizable, and not all sinks are dead ends: some findings of routine research are used in further research even when they pose no new problems. Most data are of this kind.) Routine research is closed, original research open ended. The former is straight, the latter tortuous: it rarely attains its goal in a direct manner, usually it zig-zags, sometimes wanders off, and at other times derails. What one reads in the published reports is an expurgated and embellished version—which is just as well, for a faithful chronicle would be too long, convoluted, incoherent, full of uninteresting detail; and most mistakes are unenlightening except to those who make them.

(Routine research can be construed as a partial function from problems to sets of findings, or $R:\Pi \to 2^F$, where F is the set of all old and new, actual and potential, outcomes of research in the same field as Π. F does not include the null finding 0. We say that research into problem $\pi \in \Pi$ ensues in a knowledge gain $R(\pi) = f \in F$ provided $f \neq 0$. In the case of routine research the background knowledge, the methods and the aim of research are fixed and taken for granted, so there is no need to include them in the domain of R. On the other hand original research can vary these items, and results not only in findings but also in further problems. Therefore it may be construed as a partial function from quadruples $\langle$bits of knowledge, problem, method, aim$\rangle$ to couples $\langle$set of findings, problem$\rangle$, or $R:2^B \times \Pi \times M \times A \to 2^F \times \Pi$. See Figure 7.6.)

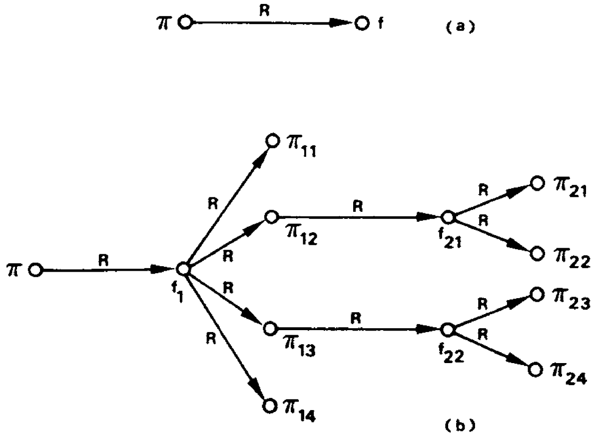


Fig. 7.6. (*a*) Routine research transforms problems into findings. (*b*) Original research transforms problems into findings or into further problems. Some problems are not investigated at all.

(The two functions, routine research and original research, may be thought to be embodied in sequential machines. Routine problem solvers may be construed as sequential machines with fixed next state and output functions: every input takes them to another state in a fixed state space, and induces an output from a fixed set of possible outputs. On the other hand original problem solvers may be regarded as probabilistic and self-organizing sequential machines, i.e. machines with variable next state and output functions, and moreover probabilistic functions. However, such devices are not machines proper, for there is no way to design or program a machine to enlarge its own state space, or create its own next state and output functions, but to specification—in which case it is a routine not an original problem solver.)

Our description of the general strategy for handling research problems tells us *what* is to be done but not *how* it is to be done. The how is a matter for *heuristics*, or the art of problem handling. The following heuristic suggestions may help go through the research problem solving process. (See Pólya, 1954, 1957; Bunge 1967a.)

1. *Search for a good problem*

Scan the background knowledge for a gap that you feel is worth filling and you might fill yourself.

Consult with coworkers or assistants, colleagues or teachers.

Pay no attention to pessimistic advice if you feel sure you have got hold of a good problem and can afford to fail in the attempt to solve it.

2. *Choose the right approach*

Start by locating the problem in a discipline or a group of disciplines.

Adopt an approach—basic or applied, atomistic or systemic, etc.

Determine the minimal background knowledge required to handle the problem.

Find out the recent history of the problem or that of its predecessors.

3. *Locate the problem*

Determine whether the problem is one of original research or routine, substantive or methodological, conceptual or empirical.

4. *State the problem clearly*

Minimize the vagueness of concepts and the ambiguity of signs.

Select adequate symbols—as simple, suggestive, and standard as possible.

Draw sketches or diagrams, build preliminary scale models, or make preliminary calculations.

5. *Make a further study of the problem*

Identify the premises and the unknowns.

Unearth the most important presuppositions and examine any that look suspicious. (As a result you may decide to switch to another problem, e.g. the problem of reconstructing such presuppositions.)

Eliminate the redundant information ("noise").

Compress and simplify the data.

Formulate and organize your preliminary hypotheses and rules of procedure.

Introduce simplifying assumptions or devices.

Look for similar solved problems or subproblems (components of the given problem).

Vary the constituents or the formulation of the problem, in an attempt to transform the given problem into a homologous problem in an adjoining field.

6. *Design a research project (plan, program)*

Reexamine and reorganize your preliminary hypotheses and rules of procedure.

Organize the steps of the proposed research process according to their logical priority (A logically presupposes B), epistemological priority (Knowledge of A presupposes that of B), or pragmatic priority (Doing A requires doing B first).

Choose the technique adequate to the nature of the problem and the kind of solution wanted.

Estimate the possible advantages and shortcomings of the various available techniques, if there are any.

In case no technique is available, state the strategic problem of devising one, and attack this one first.

Choose the materials—drugs, animals, locations, etc.—and the instruments—centrifuges, thermometers, computers, etc.—likely to be used.

Choose the coworkers, assistants, and technicians.

Secure the financial means.

Make provisions for the unexpected—in particular for changes in plan or even in problem.

Be prepared to have your project rejected for being either too poor or too good.

The preceding operations constitute the preliminary stage of the problem-solving process. Although they do not solve the problem, they do prepare it for attack, and they gather all the means likely to be required to solve it—or to show that the problem has no solution with such means. Actually the preliminary stage is usually the most difficult of all: a correct statement and a good strategy is half the solution. No wonder then that such preparation may consume as much time and ingenuity as the implementation of the research project.

The question of formulation is so important that we must insist on it, the more so since most problems in everyday life can be solved without formulating them explicitly, for we have learned well-tried routines to solve them. Other problems must be formulated explicitly if only because they admit of alternative formulations and therefore alternative solutions. (Better: a raw problem is actually a whole set of well formulated problems.) For example, if we wish to find out where two curves meet, we must either draw the curves (obtaining an approximate solution) or solve a certain equation. If we wish to find out whether a given material conducts electricity, we may insert it in an electric circuit and touch it (at our peril) or watch whether the pointer of the ammeter moves—or make assumptions about the composition and structure of the material, and compute its conductivity with the help of quantum mechanics. In science, technology and the humanities all problems have to be formulated explicitly and with maximum clarity. Such formulation (or refinement of the raw initial

problem) is seldom unique: it depends on the state of the art as well as on the outlook, experience, talent, and aims of the investigator. (For the importance of such personal factors see Polanyi (1958) and Holton (1973).)

The preliminary study of a problem culminates in a *research project*, plan or program. Any such project looks, in a nutshell, like this: Given a certain background knowledge b, and an approach a, attack problem π on the basis of programmatic hypothesis h, aiming at goal g, and using means (among them methods) m (available or to be produced). In short, a research project is a sextuple $p = \langle b, a, \pi, h, g, m \rangle$. *Example 1*: Given the available body of sociological knowledge, and ordinary mathematics, and adopting the systemic approach, build a comprehensive theory of social relations, on the basis of the programmatic hypothesis that all social relations are either interpersonal or intersystemic; the goal is to explain social behavior, and the means the standard method (or style) of model construction in the "hard" sciences. *Example 2*: Given the available body of neurophysiological and psychological knowledge, and adopting the bottom-up approach, identify the neuronal systems involved in thinking abstract ideas, on the basis of the programmatic hypothesis that thinking is the specific function of certain plastic neuronal systems; the goal is to furnish a deep explanation of thinking, and the means all the paraphernalia of neuroscience, experimental psychology, and mathematics.

Every independent investigator has some research project or other, whether original with him or supplied by somebody else—laboratory director, thesis advisor, senior colleague, etc. If vast and ambitious, a research plan is likely to be shared by, or passed on to, other investigators. And, on important issues, several competing research projects may be at work contemporaneously. Hence every field of active inquiry may be viewed as a changing bundle of more or less interdependent research projects in progress. (There is no harm in duplication: on the contrary, duplication is necessary for control.)

The way to become an investigator, i.e. to master the craft of research, is not by becoming acquainted with instances of good science—e.g. by reading history of science—or with methodology—e.g. by reading this book. Both help but neither is a substitute for participating in a research project at the knee of some master—much the same way that craftsmen and artists learned their trade in the Middle Ages and the Renaissance. The reason is that learning to do research is not just acquiring some information but rather acquiring certain habits, attitudes, and aspirations—all of which are a matter of know-how.

General methodology cannot go much further into the matter of handling problems. Any more specific suggestions would take us into specific branches of research. For example, a useful heuristic rule in applied mathematics is: "If a term of the form $A_x B_y$ occurs in a calculation, try adding $-A_y B_x$ to it, for the result is the z-component of the vector product of A and B". Biologists use Cuvier's rule: "If two organisms (or fossils thereof) exhibit two homologous traits, look for more". Historians make use (and often abuse) of the rule: "Look for precursors of every novelty". And doctrinaires abide by the rule: "Distrust all new ideas or behavior patterns". There are hundreds of special heuristic rules of this kind, which methodology has yet to gather and study.

4. Analysis of problems

4.1. *Logic and Semantics of Problems*

We proceed now to a logical and semantical analysis of problems. Ordinary logic and semantics do not account for them because they are not propositions and therefore can be neither true nor false. Nor can linguistics help, for problems can be expressed either by interrogatives or by commands or even by sentences. Thus 'What was that?', 'Find out what that was!', and 'You are being told to find out what that was' express the same problem. The only reasonable approach to the analysis of problems is epistemological, i.e. to regard them as the starters of inquiry processes. This is the approach adopted in an earlier work (Bunge, 1967a, Ch. 4), which we shall now review and complete.

To begin with, problems do not come in isolation but as components of approaches. Recall that we defined an *approach* as a quadruple $\langle$*Background knowledge, Problematics, Aims, Methodics*$\rangle$ (Section 3.1). But of course not every problem stands out against the entire background knowledge: only a part of the latter is required to pose and solve a problem. We call this part the *setting* or *context* of the problem. More precisely, the setting or context of a problem in a given approach is the collection of components of this approach that are related in some way (logically, semantically or pragmatically) to the concepts occurring in the formulation of the problem. For example, one suitable setting for investigating the mechanism of problem finding and solving is neuropsychology; another is the sociology of knowledge; a third is the history of knowledge.

In principle every problem can be approached in more than one manner,

and so it has more than one setting. Thus the problem of how to eliminate poverty belongs to, or arises in, ethics, ideology, politics, and applied social science. Because a problem may belong to different contexts, it may receive different solutions. For example, drug abuse is seen in a biological setting as an instance of addiction, in a medical setting as a matter of cure, in a legal setting as one of legislation and law enforcement, and so on. Each setting suggests the way to handle the problem.

Although a problem can be approached in alternative ways, these are not always equally suitable. In some cases one of the approaches is preferable to the others, and in other cases the merging of several approaches is desirable. For example, the problem of the origin of life is approached differently in science and in religion. The former approach invites making conjectures and experiments on the self-assembly of biomolecules, organelles, and eventually cells; on the other hand in many religious approaches the matter is one of faith in some dogma or legend about a divine *fiat*. In this case, then, the scientific approach will be preferred only by those who prefer truth by research to truth by revelation. In other cases different approaches are equally legitimate but partial, and must therefore be joined. Any social problem, such as poverty, ignorance, or delinquency, is a case in point: none of them can be solved within a single setting, but every one of them calls for interdisciplinary studies and measures.

In some cases the context of a problem must be shifted. Thus when a mathematician is summoned to help with a scientific or technological problem, he treats it as a problem in pure mathematics, i.e. he drops the factual interpretation of the concepts involved. For example, time will be for him just a real variable, and an equation of motion just one more equation to be investigated on its own mathematical merits. Likewise the biochemist moves biological problems over to chemistry, by reducing them to chemical reactions and studying them *in vitro*. A similar shift of context occurs when science is applied. For example, the phytotechnician must "translate" the results of the genetics laboratory into cultivation methods, crop yields, and so on. All these are cases of problem relocation or *recontextualization*.

In sum, every problem arises in some context or other—or, equivalently, it is a component of some approach or other. And it can be moved to some other context, not necessarily to advantage. There are no problems out of context, but the various possible approaches to a problem may not be equally suitable for a given purpose. And sometimes a systemic approach, involving the joining of a number of contexts, is required.

Our next task is to analyze the internal structure of problems. We postulate that every problem has three constituents: presupposition(s), generator, and solution (Bunge, 1967a, Ch. 4). Consider the problem "What thinks?", i.e. "What does the thinking?". It presupposes (takes for granted, does not question, assumes the premise of) the existence of some entity, be it the brain or the mind, that does the thinking. It is generated by the predicate "thinks", for the unknown to be uncovered occurs in the underlying formula "x thinks", or "Tx" for short. And the problem induces research resulting in a solution of the form "b thinks", or "Tb", where 'b' names a definite individual—someone's brain or mind. In other words our problem is "Which is the x such that x thinks?", which we symbolize '$(?x)Tx$'. The presupposition of this problem is "Something thinks", or "$(\exists x)Tx$", the generator is "T", and the solution "Tb", or simply "b". Logically, then, we have the sequence: *presupposition–generator–solution.* We may identify each class of problems with such a sequence.

In general, every problem is posed against a certain background of antecedent knowledge. This background contains the context or setting of the problem, and in particular its presuppositions. The latter are the propositions (data or hypotheses) that are somehow involved though not questioned in the inquiry prompted by the problem, although they may become the subject of alternative inquiries. The problem may be regarded as generated by a set of open formulas (what used to be called 'propositional functions') exhibiting gaps in knowledge, i.e. unknowns, that are to be filled by research. The formulation of the problem results, formally, from applying the operator "?" one or more times to such formulas. However, "?" is an epistemological operator, not a logical one: it means "search for" without even intimating how to conduct the search. Finally, research on a soluble problem yields eventually another set of constructs—the solution to the problem—which, when substituted into the generator, turn the latter into a set of statements that can be assigned a definite truth value.

The notion of presupposition of a problem occurs in the following definition of the important concept of meaningfulness of problems. Let C be a context included in a given body B of background or antecedent knowledge. We stipulate that problem π is *well conceived* in C only when the following conditions are jointly fulfilled: (i) π belongs to some approach $\langle B, P, A, M\rangle$ including (actual or possible) methods in M capable of handling π; (ii) all the concepts occurring in π occur also in C, hence in B,

and (iii) the presuppositions of π are compatible with C. (We require the latter to rule out crackpot questions, and do not require compatibility with the entire background B, to allow for revolutionary questions.) Otherwise π is *ill conceived* (or *makes no sense*) in C. Thus the problem "When did God create man?" is ill conceived in modern biology, which does not contain any concept of divinity, and asserts that man was not created but evolved from ape-like ancestors. In general, theological questions are meaningless in the context of science and conversely. (More in Ch. 14, Section 4.2.)

At first sight such statements as "Where is p?", "When does p happen?", "Is p true?" and "Is p valuable?" do not fit the above schema, for they seem to be generated by p itself, which is now a thing, now an event, now a proposition, now any object. Yet such problems, when restated explicitly. do prove to fit our schema. Thus "Where is p?" is best reformulated as "What are the values of the coordinates of p relative to such and such reference frame?", or $(?v)(V(p) = v)$ for short, where in this case v is a triple of real numbers. Likewise "Is p true?" is best construed as "What is the truth value of p relative to such and such body of knowledge?", or $(?v)(V(p) = v)$, where now v is any real number included in the interval $[0, 1]$. Every one of the above problems presupposes something or other—e.g. that every thing is at some place (relative to some frame), or that every proposition can be assigned a truth value (relative to some body of knowledge). And they are all questions about the value of some individual variable or other, even if the variable is hidden at first sight.

Not all problems concern individual variables or unknowns: in some cases the blanks to be filled are predicate variables. For example, the problem "What is b?" directs us to search for the cluster of properties P, so far unknown, possessed by b. This problem may be symbolized by '$(?P)Pb$', which points to the unknown P, a conjunction of attributes. Questions asking for the value(s) of one or more individual variables shall be called *individual problems*, whereas questions asking for the value(s) of one or more predicate variables shall be called *functional problems*. We assume that every elementary problem is of either of these kinds, and every complex problem a combination of problems of either kind. We shall deal below with simple and complex problems.

What about problems of the form "Does b have the property B?", in which no variable is in sight? The variable is hidden: it is the truth value of the proposition "b has the property B". In other words, the given problem is actually of the form $(?v)(V(Bb) = v)$, where V is a truth valuation function. The problem consists in finding out the precise value of v.

Similarly, all questions concerning existence can be regarded as problems about the truth values of the corresponding existence statements. Thus "Are there free quarks?" can be restated as "Is it true that there are free quarks?", which in turn is rendered explicit by "What is the truth value of the proposition that there are free quarks?". Similarly with questions concerning universality. For example, "Does every thing change?" is equivalent to "Is it true that every thing changes?", which in turn may be restated as "What is the truth value of the thesis that every thing changes?". The moral we extract from examining the above cases, characterized by the seeming absence of variables, is this: Do not allow language to mislead you—*cherchez la variable*. Note also that the question mark bears always on some unknown or variable of either the individual or the predicate type. (Furthermore "?" does not bind the variable on which it acts. Only the solution should be free from unbound variables, i.e. it should be of the general from Bb, where B is an n-ary predicate and b an n-tuple of individuals.)

(The question mark, which so far we have left undefined, can be partially defined as follows. We start by abstracting from the variable type, writing an arbitrary elementary problem in the form: $\pi(v)=(?v)G(v)$, where v is the only variable and G the generator. Our new formula subsumes both elementary problems of the individual type, i.e. $(?x)Bx$ with a fixed B, and elementary problems of the functional type, i.e. $(?X)Xb$, where b is a fixed individual. Although these problems are very different, they are equivalent in the sense that they have the same solution, namely Bb. Further, let $R:\Pi\rightarrow 2^F$ be the routine research function introduced in Section 3.2, that represents an inquiry process as a mapping from problems in Π to sets of findings in F. We stipulate the partial implicit definition

$$R(\pi)=R[(?v)G(v)]=Bb,$$

i.e. research on a problem yields the solution to the latter.)

(The generalization to a research or original problem is straightforward. In this case, as we saw in Section 3.2, the outcome of research is in general an ordered couple $\langle$Finding, New problem$\rangle$—where the finding is occasionally that the problem is not soluble with the indicated means. The case where no new problem arises is $\langle$Finding, Null problem$\rangle$, where *Null problem* $=(?x)Ox$ or $(?X)Xo$, and in turn O is the null predicate—attributable to no individual—and o the null individual—to which no predicate except O is attributable. Likewise, if the only result is a new

problem, we write: $\langle$Null finding, New problem$\rangle$, where the null finding is the one not in 2^F.)

The individual/functional dichotomy of problems cuts across many other classifications. For example, Aristotle distinguished between *what*-problems, or questions of fact, and *whether*-problems, or dialectical (logical) problems; and Pólya (1957) between *problems to find* and *problems to prove*. We shall not enter into this matter but shall comment instead on four kinds of problem that are of particular philosophical interest: inverse, why, what-if, and existence problems.

The *inverse* or *dual* of a problem is one where the data of the latter are the unknowns of the former and conversely. Thus if computing a value of a function is a direct problem, then calculating the argument(s) from the value is the corresponding inverse problem. Here is a small random sample of pairs direct-inverse problems. Given an equation, find its solutions; given a function, find the equation it solves. Given a theory, check a theorem in it; given the latter, find whether it is a theorem of a given theory (i.e. solve the completeness problem for the latter). Given the mass of a particle and the force(s) acting on it, find the particle trajectory; given the mass and the trajectory, find the force(s).

The last mentioned example illustrates a very common situation in all fields of inquiry, that of analysis and synthesis. The *analysis* problem is: Given a system (i.e. knowing or assuming its composition, environment, and structure), find its behavior. The inverse problem is that of *synthesis*: Given a behavior, find or design a system that realizes it. For example, find a kind of molecule or material, machine or cultivar, social system or teaching method, that will behave in a prescribed way or achieve given goals. Typically, the problems of technology and of social action are synthesis problems; with luck they are the inverses of scientific problems with known solutions. As a rule inverse problems are far harder than the corresponding direct problems, if only because they usually have multiple solutions.

Thus think of the following problem pairs. (*a*) Direct problem: Given a body, a screen, and a projection method, find the projection or shadow of the body on the screen; inverse problem: given the projection on the screen (e.g. a photograph) and the projection method, reconstruct the body. (*b*) Direct problem: Given a social policy, find the state of the society resulting from its application; inverse problem: given a desirable state of society, find a social policy likely to achieve it. (*c*) Direct problem: $(?x)(\sin b = x)$; inverse problem: $(?x)(\sin x = b)$ or, in an equivalent

notation, $(?x)(sin^{-1} b = x)$. In this case the direct problem has a single-membered solution, for *sin* is a function; but, because *sin* is not a 1:1 function, the inverse problem has a multiple—in fact an infinitely-membered solution. (If a is a solution to it, so is $a + 2n\pi$, where n is an arbitrary integer.) So much for the fascinating and ill-studied problem of inverse problems.

Why-problems are of the form "$(?p)\ (p \vdash q)$", i.e. "What premises p entail the given q?". This is the kind of problem that the theory builder faces: his givens are low level generalizations, and his desideratum is to invent high level generalizations entailing the low level ones. (More in Ch. 10.) *Why*-problems are the inverses of *what-if* problems, which have the form $(?q)\ (p \vdash q)$, or "What are the consequences of p?", or "What would happen if p were the case?". These are dialectical problems, or problems to prove. What-if questions fall into two kinds: those with false and those with true presuppositions. "What if we could live without eating?" illustrates the former kind. Such questions generate fiction rather than knowledge. On the other hand a problem such as "What if the nebular red shift were due to some cause other than the expansion of the universe?" violates no known law of physics and is fruitful, for it suggests inquiring into the possibility of a cosmology without the Big-Bang.

Existence questions are among the most difficult of all. They can be formulated thus: "Are there objects of kind K?", or, more generally, "Are objects of kind K possible?" where K is determined by fairly exact predicates. That question can in turn be restated as "What is the truth value of the statement that K is or could be nonempty?". They are of two main types: formal (e.g. existence of solutions to a given equation) and factual (e.g. existence of a planet). Consequently they are to be investigated conceptually in the former case, and empirically as well as conceptually in the second. (Recall that the expression 'exists' is ambiguous.) In science and technology existence (or possibility) problems are not whimsical: there is always some ground to conjecture that something exists or may be brought into existence. And answering such question is prior to asking the more specific question 'How do K's exist?', or 'How are K's possible?'.

Factual existence problems can in turn be classed in various ways, e.g. thus: (*a*) the most general ontological existence problem: 'What kinds of thing are there (really)?'; (*b*) scientific existence problems, of the form "Does the such and such exist?"; (*c*) the general epistemological existence problem, "To what may we legitimately attribute existence?"; (*d*) the most general technological existence problem, "What kinds of thing can we

design and bring into existence?"; (*e*) special technological problems, of the form "Can we design and construct such and such things?"; (*f*) the most general ethical existence problem, "What ought we to bring into existence?"; and (*g*) special ethical existence problems, of the form "Ought we to bring such and such into existence?".

The ontological question is wider than the partial scientific existence problems. (It would seem that there is a prior ontological problem to be solved, namely "Why is there something rather than nothing?" However, this problem makes sense only in a non-naturalistic context, for it presupposes that there is a Creator and that it might have been possible for Him not to go to the trouble of making the world. The problem makes no sense in a naturalistic ontology, let alone in science.) The epistemological question has a deceptively simple answer: We can attribute existence to whatever exists really. This answer is unhelpful, so we must settle for a weaker one, namely: We may attribute real existence to whatever we have good empirical, and if possible also theoretical, grounds to conjecture that it exists. (For a definition of real existence see Bunge (1981a).) The general technological question is of course broader than any specific technological problem such as "Are there antidotes to such and such poison?" or "How is that subject best taught to undergraduates?". The same holds for the relation between the general and the special ethical questions. But what is such a relation between a general and a special problem, or between a problem and its subproblems if any? Let us find out.

Consider any two problems. We can form a third problem out of them if we set up the task of solving one *or* the other, and a fourth problem if we propose to solve the one *and* the other. We say that in the first case our compound problem is *disjunctive*, in the second *conjunctive*. The solution to a disjunctive problem is the disjunction of the solutions to its components, whereas the solution to a conjunctive problem is the conjunction of the solutions to its components. And we form the *negate* of a problem if we negate its generator. The solution to the negate of a problem is the negation of the solution to the problem.

The three operations can be combined in a number of ways to form problems of any complexity. Conversely, a complex problem can be analyzed in terms of the disjunction, conjunction, or negation of elementary problems. For example, the algebraic problem $(?x)$ (x is a complex number & $x^2 + x - 2 = 0$) can be rewritten: $(?x)\,[\mathbb{C}x \,\&\, (x-1)(x+2) = 0]$, and thus be decomposed into two elementary problems, namely $(?x)$ $(\mathbb{C}x \,\&\, x - 1 = 0)$ and $(?x)$ $(\mathbb{C}x \,\&\, x + 2 = 0)$. The original problem is thus

seen to be of the disjunctive type, and its solution: $(a=1) \vee (b=-2)$. Whereas sometimes it pays to analyze a complex problem into its constituents, at other times it is advantageous to switch from a problem to its negate. Thus to find out whether a given person is celibate we may assume that she is married, and search for her spouse; likewise to find out whether a thing is far away we had better start by exploring the neighborhood.

(The three operations on problems which we have just introduced may be symbolized as follows. Let $\pi(v) = (?v)Gv$ be a problem with generator G and solution $G(a)$. Then the negate of the problem is defined to be

$$-\pi =_{df} \neg G.$$

Call now $\pi(v_1) = (?v_1)G_1(v_1)$ and $\pi(v_2) = (?v_2)G(v_2)$ two problems with solutions $G_1(a)$ and $G_2(b)$ respectively. Then the disjunction and the conjunction of these problems are defined as

$$\pi(v_1) \curlyvee \pi(v_2) =_{df} [G(a,b) = G_1(a) \vee G_2(b)],$$

$$\pi(v_1) \curlywedge \pi(v_2) =_{df} [G(a,b) = G_1(a) \& G_2(b)],$$

respectively, where $G(a, b)$ is the solution to the corresponding binary problem.)

On the basis of the three operations on problems, which we have introduced, more complex problem forms can be produced and analyzed. Also, the very concept of a complex problem can be defined, namely thus: A *complex* (or *compound* or *molecular*) problem is one where either the negation or the conjunction or the disjunction of problems occurs. The definition of the dual concept is obvious: a problem is *elementary* (or *simple* or *atomic*) if it is not complex. Sometimes it pays to increase the complexity of a problem, as when an insoluble problem about real functions is translated into a soluble one about complex functions, and when an economic problem is regarded as just one component of a complex social problem. However, most of the time we wish to simplify problems by analyzing them into disjuncts, conjuncts, or negates. Thus a triatomic problem may be analyzed into a disjunction of conjunctions. (In effecting such analyses we assume the associativity and commutativity laws; we also assume that disjunction distributes over conjunction and conversely.)

Besides the three operations on problems, we introduce three useful relations between problems: those of implication, equivalence, and entailment. We stipulate that a problems *implies* a second problem if the

generator of the first implies that of the second, and that two problems are *equivalent* if their generators are. We also stipulate that a problem *entails* another problem if the generator of the former entails that of the second, i.e. if the solution to the second follows from that to the first. Finally, we say that a problem is a *subproblem* (or *special case*) of another problem if the second implies or, in particular, entails the first. For example, every homogeneous equation is a subproblem of any number of inhomogeneous equations, and the problem of minimizing alienation is a subproblem of that of maximizing participation.

We can now lay down rules for the proper formulation of problems:

Rule 1: A problem π is *well formulated* (or *proper*, or *sound*) in a context C iff

(i) π is well conceived (or makes sense) in C (see the definition of this concept at the beginning of this section);

(ii) the generator of π contains as many variables and question marks as unknowns—i.e. the statement of the problem indicates explicitly what is to be sought;

(iii) π is elementary (or atomic) of the individual or the functional type, or is a combination (by means of the operations of disjunction, conjunction, or negation) of elementary problems.

A well formulated problem is determinate or well-defined: if soluble at all with the given means it will have a unique (though possibly many-membered) solution. And, by displaying all the relevant items (the data and unknowns), as well as their relations, it should suggest the type of inquiry needed to solve it. For example, a system of m linear equations in n variables (unknowns) is *determined* if and only if $m = n$, *underdetermined* if $m < n$, and *overdetermined* if $m > n$. The problem is well formulated only in the first case. If it is underdetermined, it has no unambiguous solutions—or, if preferred, it has too many solutions. If it is overdetermined it has no solutions at all (unless some of the equations are redundant). Therefore if the problem is underdetermined we must try and get further equations or eliminate some variables. And if the problem is overdetermined we must try and reduce it to a determinate problem by eliminating $m - n$ equations.

Rule 1 should be helpful to weed out ill-formulated problems but it cannot be expected to lead to formulating good problems in the sense of Section 1.2. It does not even guarantee the formulation of well conceived problems, because it is always difficult to unearth and examine all the relevant presuppositions. Even in a highly formalized theory only those

presuppositions that have been identified by the theorist are listed. Except in trivial mathematical cases, every list of presuppositions is likely to be incomplete. Although this incompleteness causes sometimes the problem to be ill-conceived, the very discovery of such incompleteness may be the source of interesting new problems.

So far we have dealt with isolated problems. But in ordinary life and in research fields problems come in clusters, not in isolation. This is particularly true of original research in science and technology, where every problem has some precursor and triggers research ending up in further problems. (This is what is meant by the popular saying 'Everything leads to something else'.) Thus "What is it like?" precedes both "How did it evolve?" and "What is it for?". The precedence in question is epistemological in the case of cognitive problems: we must solve certain problems before being able to attack others. And it is pragmatical in the case of practical problems: we must make certain means available before we may endeavor to attain certain ends.

So, every problem, whether cognitive or practical, belongs to some *branching sequence* of problems, i.e. a partially ordered set. Discovering, planning or manipulating such partial orderings of problems is part of the strategy of research or of action. Such strategy must be sketched, if only to be modified whenever necessary, unless the search or the action are to be haphazard and therefore inefficient. Evidently, the ordering relation is this: a problem *precedes* another if solving the latter presupposes solving the former—i.e. if the solution to the first is used to solve the second.

(In the case of cognitive problems this relation can be elucidated thus. Let C_1 and C_2 be the contexts or settings of problems π_1 and π_2 respectively, and call $G_1(a)$ and $G_2(b)$ their respective solutions. Then $\pi_1 < \pi_2$ iff $G_1(a)$ is in C_2. In particular, if the problem is one of proving, C_2—which includes $G_1(a)$—entails $G_2(b)$. In the case of practical problems we can say that $\pi_1 < \pi_2$ iff the solution to the first problem is among the factual conditions—causes, actions, materials, or what have you—needed to bring about the solution to the second problem. Since practical problems are concerned with the modification of some concrete system or other, problem precedence amounts to this: $\pi_1 < \pi_2$ iff the system has to attain the state associated with the solution of π_1 before it can reach the state associated with the solution to π_2.)

(Actually problem clusters are richer than ordered sets: they are *networks*. In fact, given any problem we can form its negate, and given any two problems we can form their disjunction and their conjunction; besides,

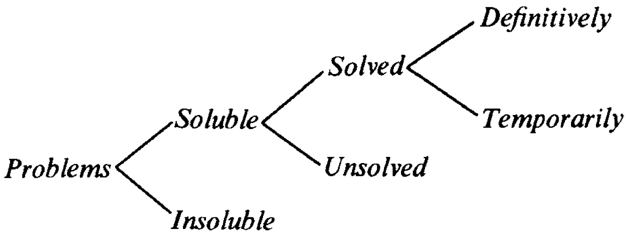
problems come in direct-inverse pairs. This suggests defining a *problem system* as a nonempty set of problems, ordered by the relation of problem precedence, together with the operation of problem disjunction, conjunction, negation, and the still unclear operation of problem inversion. The algebraic structure of such a system is yet to be investigated.)

We postulate that *every research problem belongs to some problem system*. A corollary of this assumption is that no research problem can be solved unless some other components of its system are attacked as well—some before, others at the same time. Thus the problem of synthesizing a living cell requires the solution of hundreds of problems in molecular biology, biochemistry, cell biology, evolutionary biology, instrumentation, etc. And the energy crisis cannot be solved by adopting a single measure, for it belongs to a vast problem system composed not only of technological problems but also of economic, political and cultural ones. The mere realization of the systemic character of research problems would facilitate research and expedite social readjustments.

4.2. *Solved, Unsolved, Insoluble*

Not all problems, even if well formulated, are soluble; and those which are can be solved with certain data (or assumptions) and methods but not with others. Solubility is relative to data (or assumptions) and methods rather than absolute. Thus the problem of the quadrature of the circle cannot be solved with ruler and compass but is easily solved analytically; and the problem of communicating with the deaf-mute cannot be solved using oral language but is solved by means of sign language.

In regard to their solubility problems can be classed as follows:



Except for routine problems it is impossible to prove conclusively that a given problem is soluble—short of solving it. On the other hand it is sometimes possible to prove that a problem is insoluble in a certain manner—e.g. exactly, or in a single manner—with the help of certain data

(or assumptions) and methods. For example, we cannot measure simultaneously the exact position and momentum of an electron because it does not have them; we cannot build a perpetual motion machine because energy is conserved in a closed system; and we cannot communicate with the dead because the dead happen to be dead.

All of the above are examples of total insolubility or, if preferred, of ill-conceived problems or even pseudoproblems. They are not insoluble for want of knowledge but because the world is the way it is. (For example, we cannot communicate with the dead because only the living brain is capable of thinking, and thought can be communicated only through the senses.) They are not permanent mysteries such as those of religion, which are supposed to be beyond the reach of reason and experience. Nor are they relative mysteries such as most murders, which could be solved with some more inquiry. They are just nonproblems.

The genuine but insoluble problems are those the metamathematicians call 'unsolvable': they can be *proved* to be such. For example, Gödel proved the existence of statements that are not provable within certain theories, namely all those containing arithmetic. He did not exhibit any specimens of such statements but just proved that they exist (formally), so his proof did not affect the working mathematician. Still, it undermined the faith that every mathematical problem can be solved within some theory. In factual science and in technology we cannot prove any problem to be unsolvable. Here we know that, whereas some problems have been solved, if not completely and exactly at least partially and approximately, others remain unsolved or even unstated. In factual science and technology we have no mysteries but only unsolved problems.

Here is a random sample of scientific problems still unsolved, or solved only partially, by 1983. Is the universe spatially finite or infinite? Are there black holes? What are quasars? What is the precise origin of cosmic rays? Are there reliable earthquake predictors? What caused the various glaciations? What is the function of the neutral mutations: important or insignificant? What is the function of "selfish genes" (that do not transmit hereditary traits)? Why are most molecules in organisms L-active? Does bioevolution proceed always gradually or is it punctuated by macromutations? At what point does a variety become a species? Is geographic isolation necessary for speciation? What caused the extinction of three-quarters of the biospecies at the beginning of the Tertiary era about 65 million years ago? Why do we have only half as many genes as frogs do? What is the precise lineage of our species and, in particular, whence, when

and how did the most primitive *Homo* emerge? Why do all higher vertebrates sleep? Why do most people retain their most important beliefs in the face of refutation? How do we think? How do we learn to talk? Why are some people more interested in solving problems than in nursing beliefs? What are the factors of social cohesion? Are there natural laws of the economy? Are there laws of history? How did capitalism originate? Is socialism compatible with democracy?

All of the above problems are being investigated at the time of writing, and some of them have given rise to heated controversies. Few scientists believe that there are cognitive problems both well formulated and insoluble, i.e. absolute mysteries: they leave irrationalism to theologians and philosophers. Not that it can be proved that all of the above problems are soluble: this is a matter of faith, not proof. (Solutions, not problems, are subjected to proof or disproof, conclusive or otherwise.) Yet this is no naive (uncritical and groundless) faith but one that can be justified in various manners.

First, there is the past success of inquiry every time it has been conducted scientifically and been allowed the freedom to do so. Second, there is the enormous capacity of the human brain to form and dissolve neuronal connections, to activate itself, to respond to novelty, and to seek it. Third, society needs more cognitive problem finding and solving if it is to face its most pressing problems. (Not that science can tackle practical problems, but it does produce some of the knowledge needed to solve some of them.) In short, optimism in the capacity of research is well founded. (Whether research will continue to be conducted is another question to be addressed in Ch. 13, Section 4.) Without such rational faith in research no long term research project would ever be undertaken. Radical skeptics and dogmatists have no such faith and as a consequence they never discover or invent anything but errors in others.

Our faith in inquiry, though great, is not boundless: it is tempered by the following facts. First, many solutions that seemed flawless at the time they were obtained turn out to be defective in some way or other (even in pure mathematics); consequently they have to be repaired or even abandoned. Second, some problems are given up before they are solved, either because they fall out of fashion, or because they are intractable with the available means, or because they are seen to be unimportant, or because of new unfavorable cultural or political circumstances. Third, there are limits to the information we can obtain and to the number of problems we are capable of tackling: we are not omniscient and we do not want to know

everything that can be known. (Information overload can be as bad for original research as lack of information, for it is confusing and inhibiting. And lack of information is sometimes unavoidable, particularly regarding past events: their traces have often been obliterated by other events, so there is no record to be studied.) Fourth, there are certain problems that, because of their breadth and depth, are likely to stay with us forever: they are the scientifico-philosophical and technico-philosophical problems.

Think of "What is matter?", "What is spacetime?", "What is life?", "What is mind?", "How can we know?", "What can we know?", "What is good?", "What can we do and what ought we to do?", and "What does the good society look like?". These are not mere puzzles, let alone illegitimate philosophical questions, as the pseudoanalytic counter-revolution in philosophy has labelled them. (Cf. Warnock, 1958.) Nor do they arise from a misuse of words, so they will not be solved, let alone dissolved, by linguistic analysis—the way Wittgenstein and his disciples would have us deal with philosophical problems. (Remember the famous pronouncements: "philosophical problems arise when language *goes on holiday*"—Wittgenstein, 1953, p. 19—"the clarity we are aiming at is indeed *complete* clarity. But this simply means that the philosophical problems should *completely* disappear"—*ibid.*, p. 51—and "The philosopher's treatment of a question is like the treatment of an illness"—*ibid.*, p. 91.)

Those wide scope problems are the very *raison d'être* of entire disciplines, and they are unlikely to receive final solutions even though they are being investigated and progress is being made in solving them. All grand problems are like hydras: as one head is cut another sprouts. But, unlike Hercules' hydra, they will continue to defy mankind as long as there are curious people left. Progress changes the problematics without shrinking it. On the contrary, the more we get to know, the more and tougher problems can we formulate. Progress in any epistemic field is measured not only by the problems that get solved but also by the unsolved problems that are being investigated.

As long as there are rational beings they will face and seek problems of various kinds: (*a*) *practical* problems affecting individuals and social systems; (*b*) *cognitive* problems of either the substantive or the methodological kind—or, in a different classification, empirical or theoretical; and (*c*) problems of *valuation and morals*. A correct, even if only approximate, solution to problems of any such kinds requires some knowledge, sometimes available, mostly to be sought. Moreover, the

investigation of any problem of some kind poses further problems of different kinds. Thus practical and moral problems require some substantive knowledge, and the investigation of cognitive problems requires abiding by certain methodological and moral norms. In short, the composition of each problem system is heterogeneous. The heterogeneity is sometimes such that it calls for a multidisciplinary or even an interdisciplinary approach.

5. Concluding remarks

What distinguishes modern thought from that of our ancestors 50,000 years ago is not so much better brains as the kind of problems some of us tackle and the way we approach them. Whereas our remote ancestors presumably devoted their whole brain power to solving day by day survival problems, agriculture and village life gave some people, from about 10,000 B.C. on, enough leisure to apply their wits to speculative problems, as well as to try and solve practical problems more rationally than by blind trial and error. That social revolution—the beginnings of agriculture and urbanization—caused a shift in problematics and in methodics, and this in turn gave us the type of knowledge we are proud (and sometimes ashamed) of. And, of course, some of the outcomes of such intellectual endeavors have in turn elicited profound changes in our life styles—not always for the better.

Problems are the root and the fruit of original research, be it in science, technology, or the humanities. Therefore we should correct the neglect of the very concept of a problem in the philosophy of science. More research on the logic, semantics, and heuristics of problems should be able to help us work better at finding and solving problems. In particular, it should help us avoid barking up the wrong tree and devoting our talent to pseudoproblems (such as trying to reduce social science to psychology or to biology) and *cul de sac* problems, i.e. miniproblems whose solution would make no noticeable difference to the state of knowledge or of society.

The specialist may be trusted to spot wrong solutions, but spotting wrong problems takes more than professional competence, particularly when an entire army of knights has been tilting at windmills for some time. In such cases there is a (bad) model to be imitated and even a vested interest in persisting in the effort. (The quixotic attempts to build logics of analogy and of induction, to reduce all philosophical notions to those of probability or of information, and the exact but barren investigations in

subjective probability, modal logics, and possible worlds, come to mind.) As Ackoff (1974, p. 8) puts it wisely, "We fail more often because we solve the wrong problem than because we get the wrong solution to the right problem".

A touch of epistemology might also help spot and avoid trivial problems, i.e. problems which, though well formulated, are unlikely to add significantly to human knowledge. All sciences, particularly though not exclusively the less advanced ones, are littered with trivial research, the only functions of which are to allow some individuals to make a living, and to render the volume of the literature unmanageable. Think of the human and material resources that have gone into confirming the well-known platitudes that talent depends on both genetic factors and opportunities, that groups composed of competent people perform better than bunches of incompetent people, that persons with low self-esteem are reluctant to participate in group work, that group discussions are dominated by high ability or highly vocal individuals, and so on. No wonder the announcement of such trivialities is usually greeted with the sneering question 'So what?', or (in Newyorkese) 'So what else is noo?'.

In conclusion, valuable inquiry is triggered by good problems. How does one get new problems? By reflecting critically on old problems and their solutions, as well as on new situations. And how does one get ideas to solve the new problems? By struggling with them. Problems, in sum, are the very heart of inquiry.

CHAPTER 8

CONJECTURING

Solving problems, whether practical, cognitive or moral, involves conjecturing: that this fruit is edible and that animal dangerous; that these rains announce a cold spell, and that stranger is friendly; that this stone may be suitable for fashioning an axe, and those twigs for kindling a fire; that this action may benefit my kin and that one hurt them—and so on and so forth. Every problem demands making some conjecture or other, and every conjecture poses the further problem of finding out whether it is adequate (true, efficient, or good).

We seldom face experience without some preconception and anticipation, and we never undertake scientific experimentation or technological design without some hypotheses. It is not that we are perfectly rational beings, nor that we abstain from perceiving or doing until we have formed a true theory or a suitable plan, nor that every datum is preceded by some careful cognition, but that perceiving, thinking and doing go often hand in hand. To be sure we are often taken by surprise, unprepared to react to or understand what happens to us. But only those who know and expect something can be surprised, and only those who make some conjectures about surprising events, and care to check such conjectures, can hope to survive, or at least to subsist as rational beings.

1. From preconception to hypothesis

1.1. *Psychobiology of Conjecturing*

We all know that the animals capable of learning can learn to expect certain events to follow others, and to perform certain actions after certain events occur, or in order to attain certain goals. But how do they learn: by reinforcement of the adequate expectations and actions, or by forming and trying out alternative hypotheses? Behaviorism gives the former answer, and cognitivism the latter. (Note, by the way, that this disjunctive problem contains the hypotheses we wish to discuss.) Perhaps both schools are right: maybe the two learning mechanisms operate in the higher animals, whereas the lower ones learn only, if at all, by reinforcement. But, since this

chapter happens to be about the nature of hypotheses and their role in problem solving, in this case we are likely to learn more from cognitive psychology than from behaviorism.

Classical—i.e. mentalistic and introspective—psychology had paid much attention to thinking and, in particular, to hypothesis formation, though usually without employing the experimental method. (See Claparède (1934) for an excellent detailed review.) In particular, Gestalt psychology had emphasized the role of hypothesis in problem solving among primates, but it had ignored the question of testing. The animal faced with a problem was said to start by making a few haphazard trials, then engage in serious thinking until struck by a clap of insight, upon which he acted in the correct way. There is some truth in this picture but it is incomplete and inaccurate. Preliminary exploration is often systematic (because guided by experience, by cues or by clues) rather than random. And hypotheses are tested (in imagination or empirically) no sooner they are formed. Moreover the outcome of such tests often suggests new hypotheses or poses new problems.

The breakthrough in the experimental study of the role of hypothesis in problem solving was inspired by Lashley (1929) who, in an influential book, suggested that learning proceeds by making hypotheses and putting them to the test. Krechevsky (1932) and Tolman and Krechevsky (1933) took up this idea and tested it in rats. They found that the rat does not behave haphazardly during the early stages of a problem-solving process but proceeds quite methodically, "attempting various solutions and giving them up when they fail, until he hits finally upon the 'correct' one" (Krechevsky, 1932). This work was severely attacked by the behaviorists, in particular K. Spence, and so the incipient American branch of cognitive psychology went underground. (It never ceased to be in the foreground of European psychology.)

The hypothesis that the higher vertebrates learn by making hypotheses and testing them resurfaced in America with *A Study of Thinking* by Bruner, Goodnow and Austin (1956). It is now being investigated vigorously both in the laboratory and mathematically. (Cf. Restle, 1962; Levine, 1966, 1974; Mayer, 1977.) One finding of this research is that human learners are not passive: they do not respond always in the same case to a given stimulus, but their response depends upon their experience, expectations, etc. Another is that the correct response is not strengthened gradually by reinforcement: in some learning tasks the subject can learn at a single trial. Both results, known or suspected by nonbehaviorist psychology, refuted behaviorism.

Restle (1962) has interpreted these experiments in the following way. At the outset of a learning task (discrimination learning, maze learning, or concept formation) the subject has a universe of hypotheses or strategies. He samples them first at random, and acts accordingly. If he hits on the correct hypothesis or strategy, he sticks to it without making any further errors. Otherwise he tries again by resampling. So he may choose the just rejected hypothesis with the same probability as any other one—a rather implausible assumption. (The only rationale for it seems to be that it allows its author to model the process on that of a lottery with replacement, and so with constant probability for each strategy or hypothesis.)

This model seems unsatisfactory if only because it contains the zero-memory hypothesis just commented upon. But there is more. If a problem is soluble at all, then either there are methods for tackling it or there are none. If there are, then the subject may know some of them or he may not know any. If he does, then he is likely to try first the simplest, or the one he remembers best. And if he does not know of any such method he may give up or give it a try. If the latter, he is likely to proceed by groping (*tâtonnement*) but not wholly in the dark: one is always guided by certain clues, and one can always exert one's imagination.

That hypothesis formation is seldom haphazard is suggested by observations and experiments on the way children and adults solve qualitative combinatorial problems. Take for instance anagrams, and in particular the 5 letter word 'ilrta', meaningless in English. There are 119 variations of this string of letters but only a few of them, such as 'trail' and 'trial', belong to the English vocabulary and therefore qualify as solutions. Experiment shows that, typically, subjects do not try all the 119 permutations but begin with the combinations that have the highest frequencies, e.g. 'tr' rather 'rt'. (The strength of such clues is such, that subjects find it usually difficult to break the letter pairs that occur frequently, so a problem word such as 'beach' is harder to crack than its anagram 'hbcae'.) In any case, the time it takes most subjects to solve such problems varies typically between 10 seconds and 2 minutes, and is therefore too short to allow for the examination of all the alternatives. (See, e.g. Mayer, 1977.)

So, even in the absence of methods one usually employs some clues, not always explicit, i.e. knowledge—explicit or tacit—of certain regularities that he has learned before. That is, hypotheses, even when they seem to come by sudden illumination, do not jump out of nothing but are somehow generated from available knowledge. Exactly how they are formed we do not know yet, but it is likely that thinking up a new hypothesis is identical

with the activation of a new psychon. It would also seem that there are several different modes of generation. Let us consider quickly a few of them.

To begin with there is induction from observed cases, either from very many or from just a few or even one. Thus it usually takes a young dog a single disagreeable experience with fire to form the generalization that all fires burn. Because of the universality of induction among the higher vertebrates, it cannot be the mark of science—*pace* the opinion of most inductivist philosophers. What is a (though not the) mark of science is the control or test of generalizations, inductive or not, by means of observation or experiment, as well as the attempt to explain them—e.g. to understand why fires burn. Without such tests, i.e. without going beyond the empirical sources of our hypotheses, we are likely to accumulate false beliefs, such as the time-honored ones that every existent fulfils some function in a well designed order; that all celestial bodies revolve around our planet; that every comet announces some major catastrophe; that all societies are stratified; that all conquered or colonized peoples are inferior; that all humans believe in deities; and that all knowledge comes from the senses. After all, every one of these superstitions has a good empirical basis.

At other times we form hypotheses by noting associations. Thus if we know that A and B often occur together, or one after the other, we are likely to conjecture that A and B are somehow correlated or even functionally related. (This requires the prior nontrivial isolation of A and B from the jumble of phenomena, or even the hypothesizing of their existence.) Unless we are engaged in scientific or technological research, we are likely to care only for the joint (or successive) occurrence of A and B, disregarding the three remaining cases, which are just as important, namely A and not-B, not-A and B, and not-A and not-B. For example, even experienced nurses will "see" nonexisting correlations between disease and symptom for disregarding such negative cases; and, in general, "adult subjects with no statistical training apparently have no adequate conception of correlation" (Smedslund, 1963, p. 165). The fallacy of nonexisting association is of course at the root of many superstitions, such as the beliefs in magic, astrology, and miraculous cures.

At other times we form hypotheses on the strength of similarities, real or imaginary. Thus Faraday invented the concept of electromagnetic field by analogy with the concept of an elastic body; and Darwin drew inspiration from his readings about economic competition and overpopulation. At still other times we form hypotheses with the help of general principles. Thus if

we note that the total energy of a system has decreased or increased, we use the principle of conservation of energy to conjecture that the system is not closed, so that there has been an energy leak.

However, the deepest and therefore most important hypotheses are not suggested by recurrent phenomena, repeated associations, or real analogies, but they are newly invented. The reason is that, by definition, a deep hypothesis concerns facts that are not accessible to direct observation. The phenomenal world is complex and messy. A meticulous recording of everything we perceive is unilluminating because of excessive detail. To understand what we perceive we must imagine imperceptible things or connections, or analyze in imagination what perception presents as a solid block.

Thus Harvey, who had at his fingertips just as many anatomical data as his predecessors, was the first to conjecture that the heart, the arteries and the veins form a system of closed circuits. Part of this conjecture was the hypothesis that there are capillaries, invisible to the naked eye, that join the ends of the arteries to the beginnings of the veins. This hypothesis was confirmed, with the help of the newly invented microscope, only after Harvey's death. Likewise Maxwell, guided by the general hypothesis that all electric currents are closed, postulated the existence of a displacement current between the plates of a discharging condenser. This hypothesis was confirmed after his death. And Ramón y Cajal, using Golgi's staining technique, showed that every region of the brain is composed of neurons. Although his preparations exhibited only discrete cells, he conjectured that the brain forms a system, and this hypothesis was confirmed only half a century later with the discovery of synaptic transmission. In all three cases the initial data had been insufficient even to describe what happened. In all three cases the hypotheses turned out not only to be true but also tremendously fertile. But although deep, audacious and path-breaking, all three conjectures were initially just explicit and testable educated guesses. In this regard they were no different from the thousands of hypotheses, of varying grades, that every active scientist, technologist, humanist or professional frames in the course of a year.

1.2. *Hypothesis*

In mathematics the word 'hypothesis' designates an assumption such as a datum of a problem. For example, in proving a theorem about right angle triangles we may, at a certain point, need to remind ourselves that, *by*

hypothesis, the triangles in question have a right angle. Tentative statements, such as candidates to theorems, are called 'conjectures' in mathematics. We shall be concerned throughout with conjectures.

We make guesses or conjectures all the time, in ordinary life as well as in the various sciences, technologies, arts, and professional activities. However, not all of our conjectures deserve being called 'hypotheses'. Only the *educated guesses* that are *formulated explicitly* and are *in need of testing* qualify as hypotheses. The formulation must be explicit but need not be propositional: the design of an artifact, a historical reconstruction, the reconstruction of a lost original, and the program of a course of action qualify as hypotheses. See Figure 8.1.

We have employed the expression 'educated guess' to designate a guess that, far from being wild, is compatible with some background knowledge. For example, suppose we are asked to guess the number of children in Peanutia, a Third World country with one million people. A wild guess is an arbitrary number between 0 and 1,000,000. An educated guess is this: since in most Third World countries children constitute half the population, Peanutia has about 500,000 children. This guess is educated and it can be put to the test by looking up the latest census figures: it is a testable hypothesis. The guesses of all experts are of the same kind and are formed in the same way: by combining bits of specific information (e.g. the total population of a certain Third World country) with generalizations (e.g. that in Third World nations persons up to 15 years old constitute half the population).

Some of our hypotheses concern observed or observable facts, like the conjectures we make while driving, about the congestion of traffic ahead, or the coming rainfall, or the possible outcome of a test (e.g. an experiment), or the performance of an as yet untried artifact. Other conjectures are about unobserved and perhaps unobservable facts, like those about other people's intentions we make while driving, or the behavior of our relatives while they are not being observed, or the efficiency of a new form of organization, or the nonphenomenal properties of things of any kinds.

The hypotheses concerning as yet unobserved entities or properties must be conceived, of course, before the corresponding direct or indirect observations can be performed. At the time of conceiving such hypotheses there may be data, but all they do is to pose the problem of explaining them and constraining the possible hypotheses. Moreover the hypotheses are needed in order to design observations aiming at showing the existence or nonexistence of such entities or properties. Recall some outstanding

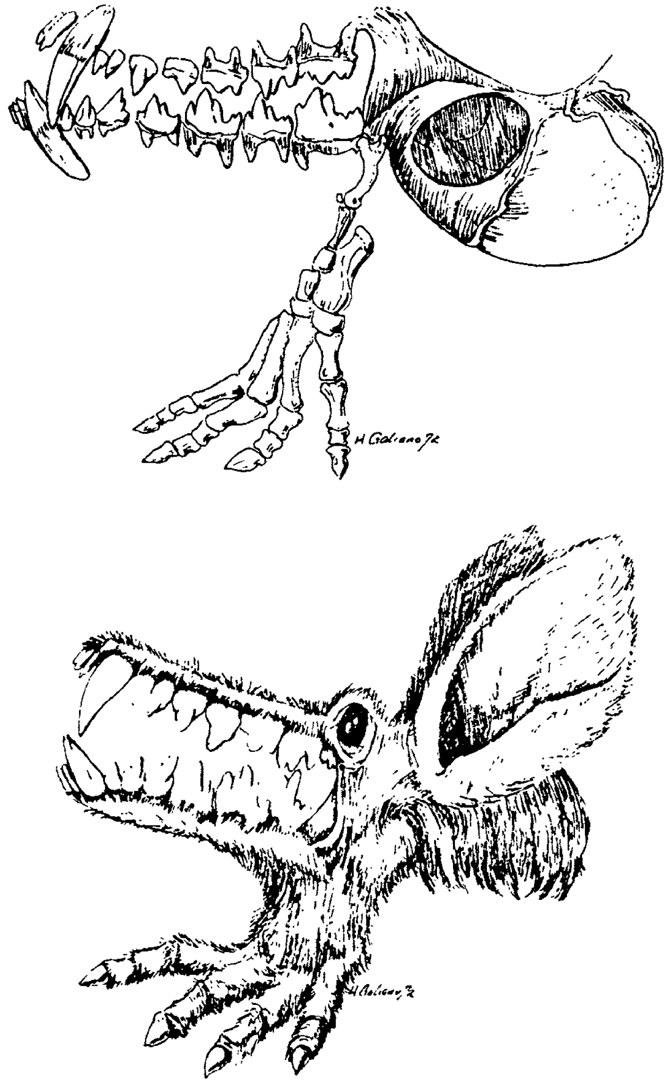


Fig. 8.1. Hypothetical reconstruction of the past: the case of fossil mammals according to McKenna (1976, pp. 246–7). (*a*) Strict adherence to the data (fossil bones) would lead us to believe, as the late A.S. Romer used to say in jest, that the extinct mammals were merely teeth that lived, gave birth directly to more teeth, and died. (*b*) We know better: the early mammals had also a foot, an auditory apparatus, and fur.

examples of this kind of hypothesis: atoms were hypothesized twenty-three centuries before their existence could be established experimentally; genes and antibodies were postulated half a century before being observed and identified chemically; nuclear, interatomic and intermolecular forces are hypothesized in order to explain the outcome of scattering experiments, the main function of which is to furnish some indications about those forces.

Some of the unobserved facts we make hypotheses about are in principle observable and, with ingenuity and luck, are eventually observed. Others are unobservable *in principle*—at least according to our present knowledge. A clear case in point is that of historical hypotheses, i.e. hypotheses about the past of the universe or of any part of it. All such hypotheses are about alleged facts that, not existing any longer, are unobservable in principle. To be sure, some such facts are of the repetitive kind and so we can use our knowledge about current facts of the same kind. For example, the emission and absorption of light can be assumed to "obey" now the same laws it "obeyed" a thousand million years ago; and we can assume that the earliest hominids had most if not all of our own basic needs.

But other facts are unique or unrepeatable. Hence if past and unobserved they must be reconstructed hypothetically on the strength of whatever traces they may have left, and of some regularities. For example, the feeding habits of animals belonging to extinct species can be reconstructed in either of the following ways. One is to analyze the undigested remains of the excreta, which are rather exceptional findings. If this method cannot be used for lack of data, the palaeontologist attempts to reconstruct the physiology of the animal on the basis of its fossil remains—an exciting and chancy exercise. A third, no less fallible way, is to extrapolate back from the known feeding habits of living relatives. The best procedure would of course be to combine all three, but this is seldom possible. And in any event each of the three methods involves making hypotheses. In general, the shorter on empirical evidence the longer on speculation, the longer on evidence the longer on well-founded hypothesis, and the longer on the latter the longer on new discoveries or inventions. To abstain from speculating altogether is to condemn oneself to living in the present without understanding it. Not speculation in itself but only wild, i.e. groundless and untestable, speculation is to be shunned. Let sound speculation flourish.

Every inquiry, even the simplest data gathering operation, involves some hypotheses about the object of study as well as on the manner of studying it. (The former are substantive, the latter methodological hypotheses.) For example, in going about their data gathering, anthropologists and field

linguists have usually assumed that the uneducated elders of the tribe or the village were the repositories of tradition and language in its purest form, before being corrupted by the radio, tourists, or visiting scientists. In this way they collected many data that were not representative of the population as a whole and even, in some cases, histories deliberately made up to please or fool the scientists. That such mishaps are difficult to avoid is beside the point. The point is that data, whether in science, technology, the humanities, or professional life, are always sought in the light of some substantive and methodological hypotheses: before rushing to make observations or experiments we must choose the kind of data we want and must decide how best to obtain them. What to do with the data once collected is another story.

According to empiricism the task of the researcher is finished once he has got his data. At most he may summarize and cautiously generalize the information thus obtained, i.e. he may formulate low level hypotheses. (A hypothesis may be said to be of the *low level* kind if it contains no substantive or technical concepts other than those occurring in the data it covers. Data-fitting curves are of this kind.) No doubt such a task is to be performed, if only because the low level hypotheses are raw material for the theorist. See Figure 8.2. But some of us want not just to know what is the case but also to understand, and this calls for inventing further hypotheses—explanatory hypotheses. In short, the process is not from data to hypotheses (the positivist schema) but from problem to hypotheses, and from these either to data (in the case of empirical research) or to further hypotheses (in the case of theoretical research), and from either to control or test—and then back to the original problem or forward to yet another problem, perhaps one generated by either the hypothesis itself or its test.

Think of any of the research processes that culminated in some of the most interesting findings of our century—e.g. the quantum theory of atoms, which is not a data package but a high level construct containing no

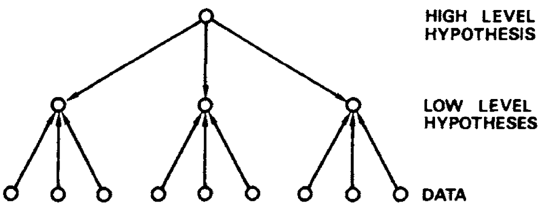


Fig. 8.2. High level hypotheses colligate and entail low level hypotheses, which in turn summarize and generalize (but do not entail) data.

data proper. The motivation for building it was the failure of classical physics to explain some facts known of old as well as some newly discovered facts. There was nothing to induce from: induction came only at the end, when evaluating the theory in the light of experimental results. Or think of the many hypotheses involved in any model of the economy of a whole nation, such as an econometric model of the US economy. Or of the many hypotheses involved in any long term development plan—or, for that matter, in any plan to turn back the social and economic clock. Or, finally, think of the history of television and the host of hypotheses of all levels, including some taken from electrodynamics and others from electron physics, involved in it.

Are there any rules or methods for conceiving hypotheses? A number of outstanding thinkers, among them Leibniz, have claimed that they had invented, or could invent, an *ars inveniendi* or logic of invention and discovery. The idea was to invent a handful of precise and comparatively simple recipes wherewith anybody could discover truths regardless of his background, talent, and motivation. Of course nobody did in fact exhibit any such universal rules, and the project died out. A few optimistic philosophers (notably Hanson, 1958) tried to revive it in our century but, predictably, they too failed. The project of inventing an *ars inveniendi* is chimaerical because originality consists in breaking new paths, not in following existing ones. It would be far less wrong-headed to investigate which chemicals, and at what concentrations, favor or inhibit original conjecturing. And it is a legitimate problem for sociologists and historians to find out what types of social condition stimulate or inhibit conjecturing. (Cultural freedom, though necessary, is insufficient. Affluence, though helpful, is neither necessary nor sufficient. A tradition of learning is ambivalent. And a philosophy contemptuous of hypotheses is inimical.)

However, there are a few methods for generating hypotheses from data. Any interpolation formula, allowing one to construct a continuous curve (a polynomial) out of a finite number of data, is such a technique. The technique for calculating linear regressions is another. These and other techniques can be incorporated into computer programs. In fact computers can be programmed to generate hypotheses of the following kinds: "A is associated with B", "A increases with B", "A decreases with B", and "C makes A and B correlated" (Hájek and Hávranek, 1978). More advanced programs, particularly DENDRAL, are employed in organic chemistry to form the plausible structure diagrams of molecules on the basis of their composition and mass spectra, the valence of their component atoms, and a

few other data and hypotheses (Lederberg and Feigenbaum, 1968). In short, it is possible to mechanize induction.

All such algorithms are of great practical value to process huge masses of data, as is the case in statistics, chemistry, medicine, and some other fields. However, it must be kept in mind that the resulting hypotheses are all *low level* ones and therefore of no great theoretical interest. In fact the hypotheses generated by algorithms out of data (*a*) colligate data (to use the term introduced by Whewell (1847)); (*b*) involve only substantive concepts occurring in the data (e.g. those of atom and valence in chemistry), and (*c*) are mathematically extremely simple-minded.

The generation of powerful hypotheses, such as differential and integral equations, is beyond the reach of any algorithm or machine program. Even simple law statements such as Snell's law of refraction cannot be generated from a bunch of data. What one can do with a computer is to *check* whether a given high level hypothesis fits a certain data base. In sum, computers can help process data and generate low level hypotheses, as well as find out how well or poorly the latter fit the data. But they cannot invent any high level hypotheses because the latter involve, by definition, concepts that fail to occur in the data.

We do not know of any rules or methods for generating high level hypotheses such as the hypothesis of natural selection, economic determinism, or the Schrödinger equation. We just grope for high level hypotheses, helped or hindered by induction or analogy, by general principles or mathematical forms, and always spurred by imagination, a knack for discerning or guessing patterns, and the philosophical conviction that there *are* patterns to be uncovered. All of the high level hypotheses are invented in an unruly way and perhaps also unconsciously. Only their precise formulation and checking must be methodical and fully conscious.

Philosophers, in sum, are no longer interested in the chimaerical project of setting up a logic of discovery or of invention. They work, on the other hand, on such problems as: (*a*) defining the very concept of a hypothesis, (*b*) classing hypotheses (e.g. into empirical and theoretical, or low level and high level, substantive and methodological, etc.); (*c*) unearthing the clues that led or misled to the formation of some hypotheses of historical interest; (*d*) establishing the conditions a hypothesis should fulfil in order to deserve being examined theoretically or empirically; (*e*) establishing the conditions for concluding that a hypothesis has been satisfactorily confirmed. These are genuine problems, and open ones for the most part. (Because of the empiricist tradition in epistemology they are often lumped

under the misnomer *Induction.*) We shall address some of them in this book.

2. Scope and depth

2.1. *Scope*

Hypotheses come in all scopes or extensions, and the real scope of a hypothesis is not always apparent but must be dug out by logical analysis. We distinguish the following kinds of hypothesis with regard to scope:

1. *Singular* hypotheses, such as "This machine should work", and "That social program should decrease unemployment".

2. *Particular*
 2.1. *Indefinite* particular hypotheses, such as "There are black holes", which specify neither place nor time and are therefore hard to refute.
 2.2. *Definite* particular hypotheses, like "There are decision-making modules in the frontal lobes of all normal higher vertebrates".

3. *General*
 3.1. *Bounded universal* hypotheses, such as "All Renaissance intellectuals were generalists", that are restricted to certain places or times.
 3.2. *Unbounded universal* hypotheses, such as the laws of physics.

Scope is no ambiguous indicator of importance. Thus the hypothesis that such and such conflict was a trade war may be more important to us, who live in the shade of world wars, than many a generalization of mirmicology or of malacology. Besides, some hypotheses are singular in one respect but universal in others. For example, the astronomical and geophysical laws concerning our planet are referentially singular but they are universal with respect to time. For example, "The Earth spins" abbreviates "For all times since its formation the Earth spins". Likewise every statistical correlation, e.g. between the height and weight of persons, is a universal statement. Indeed, "Variables A and B are strongly correlated" means that, for every value of A, the corresponding value(s) of B lie close to the linear regression line.

Nor is scope, or quantifier type, a precise indicator of degree of generality: logic is too coarse a tool of analysis as compared with mathematics. For example, "For all $x, f(x) = 0$" is a special case of "For all $x, f(x) = a$", which in turn is subsumed under "For all $x, f(x) = a + bx$",

and so on. Likewise rest is a special case of motion, uniform motion (or stationary state) a more general one, and uniform acceleration still another, even more general kind. Neither of the above generalizations is inductive, for neither consists in jumping from singulars to universals. They are generalizations from species to genera, from these to families, and so on. (Recall Ch. 6, Section 2.2.) Moreover, as a matter of historical record some of the higher order generalities were conjectured before the special cases were investigated.

Empiricists have always cautioned against the sin of overgeneralization, and have advised us to stick to singulars, the only genuine existents. There is no need for such warnings, since empirical tests will correct any excesses. Tests will also correct the dual mistake of undergeneralization, by showing that what was initially thought to hold only for a given set of facts holds also for a superset of it. (A notable recent case was quantum mechanics, originally conceived for atoms only, but later generalized to molecules, atomic nuclei, and even certain macrosystems such as superconductors.) Of the two complementary mistakes, undergeneralization is more dangerous than its dual because it may consist in not making any generalizations at all, or in stopping at low level generalizations. After all, a high level general hypothesis is better than none, for it stimulates thinking and experimenting—and it is fun.

Generality is related to inferential power, or the capacity a proposition may have of generating further propositions with the help of some formal tools. Actually even the most modest singular proposition, such as a datum, may be said to have an infinite deductive power, for it entails infinitely many conditionals where it occurs as a consequent. (I.e., e entails $h \Rightarrow e$, with h arbitrary. In fact, suppose the inference invalid, i.e. $h \Rightarrow e$ false. This holds only if h is true and e false—which contradicts the hypothesis that e is true.) Such profligacy of singulars is uselesses because, by entailing anything, they generate nothing in particular. (Inductivists, take note.) Genuine deductive power consists in the possibility of specification, in particular instantiation.

We distinguish two genera of hypothesis with regard to specifiability: the unspecifiable and the specifiable ones. Singular propositions, such as "b is an F", and particular propositions, such as "Some A's are B's" (or even "Most A's are B's"), do not entail anything definite. Likewise the outcomes of statistical analyses, such as "f percent of A's are B's", and "The standard deviation of the F distribution is such and such", entail nothing. In general statistical statements, unlike general probability statements (e.g. on

probability distributions), are deductively barren. They are terminals of inquiry and, at most, they pose the problem of setting up models for explaining them.

We distinguish two species of specifiable hypotheses: immediately and conditionally specifiable. An *immediately specifiable hypothesis* is one from which singular hypotheses can be derived by assigning definite values to variables. The empirical generalizations and the lowest level theorems of factual theories comply with this condition. For example, if a function f solves a certain ordinary differential equation, it follows that its value at point b is $f(b)$. (That this value may be difficult to compute is beside the point.) A more interesting example: the basic equations of a fundamental theory, such as classical dynamics or classical electromagnetism, are specifiable to species and to individuals by specifying the functions and parameters occurring in them—e.g. by specifying the force function, the mass value and the initial conditions in Newton's law of motion, or the four-current and the boundary values in Maxwell's laws.

A *conditionally specifiable hypothesis* is one that can be "applied" (made to refer to) individual cases, or subordinate species, upon suitable formal or semantical transformations. For instance, differential and integral equations must be solved and assigned factual interpretations before they can be so "applied". Mathematics takes care of the indicated formal transformation, but it is the responsibility of the scientist or technologist to "read" the resulting formulas in terms of concrete things, their properties, and changes in the latter. That is, the mathematical formula must be adjoined meaning assumptions or correspondence rules. (Cf. Vol. 2, Ch. 7.)

Some hypotheses are so extremely general that, in conjunction with particular or even singular bits of information, they entail further general—though somewhat less general—hypotheses. For example, if a given organic molecule consists of subsystems A and B, it can be hypothesized that A and B are the precursors of the molecule (i.e. have preexisted it), on the strength of the general postulate that every complex system has arisen by the assembly of its components. This is so general a hypothesis that it belongs in the intersection of science, technology, and philosophy (Vol. 4, Ch. 6).

Scientific, technological and humanistic research are conducted with the help or hindrance of a number of philosophical hypotheses. We mentioned a few such hypotheses of an epistemological kind in the Introduction, Section 6, as well as some hypotheses of the ontological kind in Vol. 3, Introduction. These philosophical hypotheses do not occur among the

findings of scientific, technological or humanistic research, but they guide research and are controlled by it. One such hypothesis is that logic and mathematics have no factual content and therefore (*a*) can be used in any epistemic field and (*b*) are not testable by any empirical procedures—i.e. they are a priori. Another is that we can get to know more than appearances, i.e. that it is legitimate, nay mandatory, to frame and use transempirical concepts and hypotheses. A third, that the object of research is distinguishable from the knowing subject. (To be sure, the latter may influence the former, but no such influence would be possible if they were indistinguishable, i.e. one rather than two.)

Every one of these and other philosophical hypotheses is presupposed by modern research, but this does not entail that every epistemic field has a permanent philosophical basis, or that philosophy can be the ultimate arbiter. What it does entail is that there is no separation between philosophy and the other epistemic fields; consequently that neither of them can advance without the other's support and criticism. Philosophers are not equipped to approve or question any specific hypotheses in fields other than their own—unless of course they happen to be conversant with them—but they are certainly competent to approve or question the philosophical hypotheses presupposed by such specific hypotheses. Although such philosophical presuppositions of research provide no ultimate foundation, they do constitute a provisional basis to be further analyzed, systematized, supported, or corrected. They are true to the extent that they facilitate the search for truth, barren if they neither lead nor mislead, and false if they hinder such search in any way.

2.2. *Depth*

Scope is only one of the dimensions that determine the size of a hypothesis: the other is depth. We stipulate to call a hypothesis *deep* only when it contains transempirical concepts, and otherwise *superficial*. For example, the hypothesis that syphilis is only a skin disease is literally skin-deep. It was held for thousands of years and it sheltered a number of fantastic conjectures about the origin of syphilis—e.g. that it was ordained by the stars, and that it was a divine punishment for the sin of fornication. In the wake of the microbiological revolution initiated by Pasteur, the programmatic hypothesis was formulated that syphilis was caused by a microbe. Methodical investigation into this deep (transempirical) hypothesis proved successful: in 1905 *Spirocheta pallida* was discovered and

incriminated. (Cf. Fleck, 1935.) In turn, this finding posed the technological problem of getting rid of the microbe—a problem solved only with the advent of antibiotics half a century later.

Something similar is happening nowadays with regard to cancer. Surgical practice and observation have produced plenty of accurate descriptions of the outward appearance and location of cancerous tumours; clinical observation has yielded many data on the symptoms of cancer; and histological observation has completed the picture. But neither of these procedures could possibly tell us what the mechanism of cell proliferation can be, and consequently neither of them can give us a clue to the prevention or cure of cancer. On the other hand cell biologists, molecular biologists and virologists are on the right track and are learning more and more about such mechanism. The ultimate goal is a detailed and precise (if possible mathematical) model of cell proliferation—an explanatory, not just a descriptive one. But we will not get such a system of deep hypotheses if we continue to believe that knowledge can be gained only through observation or practice; if researchers continue to adhere to the empiricist epistemology, which shuns transempirical concepts and therefore deep hypotheses; and if most science administrators continue to look upon the problem as one in medicine rather than in biology.

According to our definition, a hypothesis qualifies as deep if it contains at least one transempirical concept—aside of course from the logical or mathematical concepts occurring in it. Now, we must reckon with several degrees of depth. Thus a black box model of a system, i.e. one that ignores its internal composition and structure, is superficial, for it represents only the observable inputs and outputs. If we allow for internal states of the box, i.e. build a grey box, our representation of the system becomes somewhat deeper. Still, this depth is not impressive as long as the internal states are not identified as states of the internal mechanism of the box.

For example, the simplest model of a system is the vending machine with just two internal states, the ground state and the excited state: Figure 8.3. In the absence of external stimuli the machine stays in the ground state, i.e. it lacks initiative. (When it does exhibit initiative and starts delivering goods without being stimulated, it is judged to be out of kilter.) When a stimulus of a certain kind, e.g. a coin, impinges on the machine, the latter jumps to the excited state, produces a response of a certain kind (e.g. delivers a candy bar), and returns to its ground state. Such a system can be realized in many ways, i.e. it is compatible with a whole family of internal mechanisms—mechanical, electrical, chemical, etc.

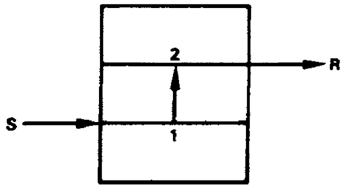


Fig. 8.3. A two state grey box model of a system. When stimulus S impinges on the system, the latter jumps from state 1 to state 2, produces response R, and then decays back to state 1. If the machine is probabilistic, it responds to the stimulus with a given probability p, so it steals dimes a fraction $1-p$ of the times it is stimulated. A simple mathematical representation of such probabilistic machine is the matrix equation $R = MS$, where $M = \begin{Vmatrix} 0 & p \\ 1 & 0 \end{Vmatrix}$ and $S = \begin{Vmatrix} 1 \\ 0 \end{Vmatrix}$, whence $R = \begin{Vmatrix} p \\ 0 \end{Vmatrix}$, with $0 < p \leq 1$.

The consumer may not mind the kind of mechanism as long as it is functional. But of course scientists and engineers are keen on discovering or designing the mechanisms of the systems they study, for they wish to understand or master them. In particular the neuropsychologist wishes to understand the neural mechanisms of cognition: he cannot remain satisfied with the "cognitivist" or "functionalist" dogma that only the "software" or program matters, the "hardware" or neural stuff being irrelevant. If a philosopher insists on remaining ignorant of such mechanism, it is his privilege. But he has no right to press for a commitment to such ignorance as the right strategy for investigating the mind. (More in Vol. 6.)

The search for, and description of, mechanisms involves *mechanismic* or *dynamical* hypotheses, i.e. deep conjectures about mechanisms of some kind—mechanical or electrical, chemical or cellular, organismic or ecological, economic or cultural. The hypotheses that gas pressure is due to molecular impacts, and that extreme class differences cause discontent, are well known examples of mechanismic conjectures. Nonmechanismic hypotheses, whether or not they contain transempirical concepts, are called *phenomenological, kinematical,* or *black box* conjectures. The thermodynamical relations between pressure, volume, and temperature, and the stimulus-response hypotheses of behavioristic learning theory, are typical examples of phenomenological conjectures.

Kinematical or black box hypotheses are easier to frame and test than their dynamical rivals. Moreover kinematical hypotheses contain adjustable parameters and can therefore be made to fit the data more perfectly than the corresponding dynamical hypotheses. It is always easier to

describe a state or a change of state than to explain how it comes about. However, we need hypotheses of both kinds for, after all, a description of facts must be available before we can proceed to explain it. Thus phenomenological hypotheses, even if they fit accurately the known data, pose the task of framing mechanismic hypotheses capable of explaining the former. We must welcome black box hypotheses, though not hail them as the culmination of the research process, and we must resist the black box philosophy that shuns mechanismic hypotheses (Bunge, 1964).

In the history of science it has been common for phenomenological and mechanismic hypotheses to conflict over a domain of facts. More often than not mechanismic hypotheses have either replaced or supplemented their phenomenological rivals. Recall the following cases: (*a*) the hypotheses about electromagnetic fields eventually displaced those of action at a distance (or direct inter-particle interaction); (*b*) chemical kinetics, which centers around equations for the rate at which chemical reactions proceed, are supplemented by hypotheses concerning reaction mechanisms (intermediate reactions), which are in turn explained by quantum chemistry in terms of forces and collisions; (*c*) the hypothesis that biopopulations evolve has been supplemented by the hypothesis that evolution proceeds by spontaneous genetic changes, natural selection, geographical isolation, etc.; (*d*) the descriptive generalizations of mentalistic and behavioristic psychology are supplemented or replaced by neuropsychological hypotheses; (*e*) the matrices of social mobility theory are supplemented by hypotheses about the economic, cultural and political "forces" causing job and status drifts.

Needless to say, a mechanismic or dynamical hypothesis is supposed to describe with precision some mechanism or other, not just state that there is, or must be, a mechanism. The latter would be a *programmatic* hypothesis inspiring the search for precise mechanismic hypotheses. Thus Maxwell's equations for the void describe in detail the mechanism of propagation of an electromagnetic field. On the other hand, when Chomsky (1980) and his school claim that the human mind contains innate mechanisms for the generation of grammatical sentences, they state only a programmatic hypothesis, for they do not even sketch such alleged mechanisms. And, since the conjecture of the inborn generative grammar is imprecise, almost any data can be construed as confirming it. Moreover, in a psychobiological perspective the mind cannot contain any mechanisms, for the mind is only a collection of neural processes (Vol. 4, Ch. 4). There are on the other hand neurophysiological mechanisms, in particular mechanisms of formation

and dissolution of plastic neural systems. In other words, the psychological mechanisms are neurophysiological. No matter, no mechanism.

What holds for psychological mechanisms holds also for social ones: they are all mechanisms of concrete things—in this case living individuals and functional social systems. For example, the hypothesis that a high rate of interest is a cause of inflation, not a deterrent to it, is true but superficial, for it does not indicate any mechanism. A more complete story, and therefore a better guide to action, is contained in the hypothesis that a high rate of interest decreases the rate of borrowing, which in turn diminishes spending, which decreases production and trade, which decreases employment, which in turn decreases spending, and so on. Incidentally, the central hypothesis of monetarism, namely that the essence of an economy is the flow of money, and therefore controlling money is the most effective way of controlling the economy, is a typically superficial (and false) hypothesis: producers and consumers, together with resources and commodities, not the means of exchange, compose an economy. (More on this in Vol. 6.)

We have declared our preference for mechanismic or dynamical hypotheses over phenomenological or kinematical ones. This preference for depth, though justified, is far from universal. In fact most people prefer black box hypotheses because they are easier to frame, fit to the data, and test. In addition, black box hypotheses have been recommended on philosophical grounds. One of them is conventionalism: If all our assumptions are mere conventions and none represents reality, why not choose the simplest, in particular the kinematical ones? Another is empiricism: If all we can get to know are appearances, why conjecture hidden mechanisms instead of remaining content with hypotheses relating the observable inputs and outputs?

Conventionalism as well as empiricism include then simplicism, or the methodological rule that recommends preferring the simplest of all hypotheses compatible with the data. There are two main variants of simplicism: moderate and radical. The former is Ockham's razor, or the principle according to which "Plurality [of entities, properties, causes, etc.] is not be be assumed without necessity". This variant of simplicism is moderate, for it discourages the invention of fictions but not that of whatever may be needed to explain things. The second variant of simplicism is summarized in the principle of economy of thought, proposed by Avenarius and Mach. According to this principle our aim as cognitive subjects should be to account for our own experience with the least possible intellectual effort. The upholders of this principle used it rather success-

fully, at the turn of the century, to ignore field physics, attack the nascent atomistics, and try to reconstruct mechanics as a kinematical theory, i.e. dispensing with the concepts of mass and force. We have learned since that simplicity is the seal of ignorance not of wisdom, and simplicism a paralyzing credo: reality is complex and accordingly our representation of it is bound to be complex. (See Planck (1910) and Bunge (1963) for criticisms of simplicism.) Still, the myth of simplicity survives among some philosophers who, even to this day, repeat Mach's dogma that a scientific theory "must provide an economical statement of the regularities which hold between observable phenomena".

3. Chance and Cause

3.1. *Probabilistic Hypotheses*

The great majority of hypotheses state associations between putatively objective properties of concrete things, such as atoms, fields, cells, persons, or social systems. There are several kinds of association and correspondingly of association hypotheses. We shall mention the following:

(1) *Qualitative association*, such as "All *A*'s are *B*'s", "Every *A* is accompanied (or preceded or followed) by some *B*", and "The occurrence of *A* prevents that of *B*".

(2) *Statistical association*, such as "*f* percent of *A*'s are *B*'s", "There is a strong positive correlation between schooling and earning", and "There is a strong negative correlation between social participation and crime".

(3) *Functional relation*, as in "Thing *a* reaches (or reached or will reach) place *b* at time *c*", and "The total consumption in any given year is a linear function of the national income in the preceding year".

(4) *Probabilistic association*, as in "The probability that *a* be at place *b* at time *c* equals *p*", and "The probability of an *A* turning into a *B* within the next time period is *p*".

(5) *Descent relation*, such as "All molecules self-assemble from their components", and "Species *A* and *B* have a common ancestor *C*".

(6) *Causal relation*, such as "Gravitation produces acceleration", and "Discontent causes rebellion".

Whereas the hypotheses of the first three types tell us (conjecturally) what goes with what, those of the last two types tell us what comes from what. The probabilistic hypotheses (4), which are a particular kind of functional relations, may represent either regular associations or transfor-

mations. The phenomenological (kinematical, black box) hypotheses are all of the types (1), (2), (3), or (4), whereas the mechanismic (dynamical) hypotheses are of the types (4), (5), or (6).

In the research process the various kinds of hypothesis remain distinct but not separate. Thus some correlations suggest or confirm descent or causal hypotheses, and the latter may result from suitably interpreting certain functional hypotheses. This statement calls for some elaboration. Take a random sample from a population, and measure two properties possessed by their members. If the two properties happen to be strongly correlated, it is legitimate (though perhaps false) to conjecture that a suitable partition of the original sample may yield a set of subsamples in each of which the correlation coefficient is maximal, i.e. equal to 1. See Figure 8.4. In other words, a correlation may hide a whole family of functional relations, and be thus just a sign of heterogeneity in the population. (Caution: the converse is not true, i.e. two variables may be functionally related without being linearly correlated. Feller (1968, I, p. 236) offers the following example. Let U and V be variables with the same distribution, hence the same mean, and let $X = U + V$ and $Y = U - V$. Then $\overline{XY} = 0$ and $\bar{Y} = 0$, whence $\rho(X, Y) = 0$. This shows that correlations do not yield general measures of dependence and that they do not generalize functions.)

In the philosophical literature (e.g. Nagel, 1961; Hempel, 1965) statistical generalizations, such as "f percent of A's are B's", and "A and B are strongly correlated", are usually mistaken for probabilistic hypotheses. One reason for this confusion is the empiricist thesis that a probability value is nothing but a long run frequency. This thesis was shown long ago to be

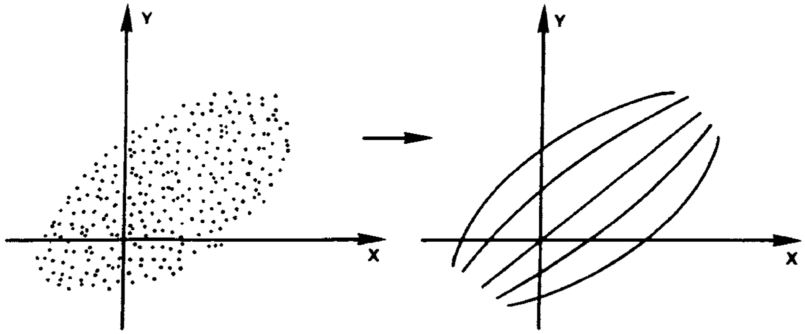


Fig. 8.4. From correlation to functional relation(s).

mathematically false (Ville, 1939). It is also philosophically untenable. For one thing, probability is a theoretical concept whereas frequency is an empirical one. For another, whereas probabilities concern potential states or events, frequencies are of actual states or events. What is true is that some (by no means all) probability statements are tested by confrontation with frequency statements. (See Bunge, 1973a, 1981c.)

Probabilistic hypotheses are theoretical and they involve, either fundamentally or derivatively, probability distributions. A good example of such hypothesis is the Maxwell–Boltzmann distribution of the velocities of the molecules in a gas: Figure 8.5. The operationalist reading of this curve is in terms of the *frequency* with which a molecule will be *found* to have a given velocity. The realist interpretation involves the concept of probability, not frequency, and says nothing about our finding anything: it says that the probability or propensity that a molecule will have a certain velocity is such and such—whether or not the velocity is measured. Considerations of measurement appear only at the test stage, when probability distributions are contrasted with frequency distributions or histograms, such as that of Figure 8.6.

A theoretical probability distribution is a very potent hypothesis, for—when conjoined with the suitable definitions—it yields any number of parameters, such as averages, moments and standard deviations (or scatters). In most applications these parameters are global or collective properties of the population concerned, be it a physical, biological, or social system. Thus the mean velocity of a gas and the average age of an animal population are global (or nondistributive) properties of the whole. However, in the quantum theory every individual entity is assigned a whole

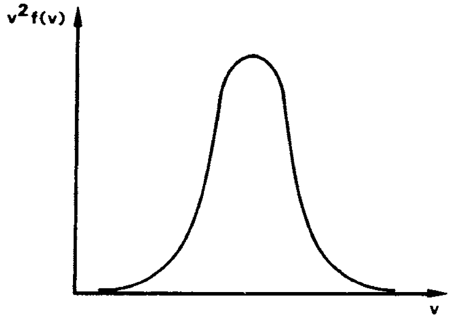


Fig. 8.5. The Maxwell–Boltzmann distribution for the probability that a molecule in a gas in equilibrium has a given velocity.

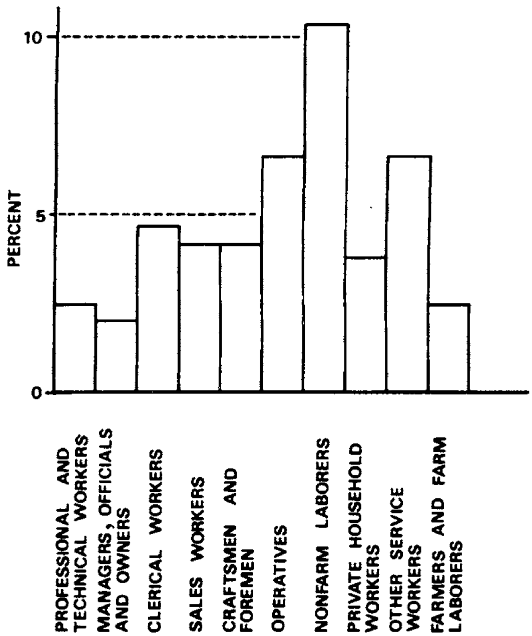


Fig. 8.6. Unemployment rates by occupation in the U.S.A. during the 70's.

set of probability distributions (of position, linear momentum, etc.). In either case some of the parameters calculated from a distribution can be checked by comparison with the data contained in a frequency distribution (histogram).

Whether classical or quantum-theoretical, probabilities are objective (though perhaps falsely posited), not subjective. (Objective probabilities are often called 'propensities', and subjective probabilities 'credences', 'degrees of rational belief', or 'personal probabilities'.) That this is so can be seen from the following considerations. First, all the probabilities occurring in science and technology concern concrete things rather than our beliefs about them. Second, in probability matters our beliefs are expressed by our favoring certain probability distributions over others—hence they are formulated as metastatements such as "I favor such and such distribution". Third, we are supposed to check such hypotheses by contrasting our probabilistic hypotheses, and their consequences, with empirical data. (For details see Bunge, 1981c.)

A probability distribution may be proposed to explain an empirical

frequency distribution or it may suggest collecting some statistic. Everyone should feel free to propose probability distributions, whether on the strength of data, or of theoretical considerations, or by sheer guessing. The only conditions are that (*a*) the hypothesis be checked, and rejected if found false, and (*b*) the thing or process in question be, or be assumed to be, random in some respect. The first condition is obvious to anyone save an extreme subjectivist. The second condition is ignored by most of the philosophers interested in probability, as shown by the popularity of the procedure of assigning probabilities to propositions such as hypotheses and data. However, no randomness, no chance for probability.

If proposed prior to the collecting of empirical data, a probability distribution is said to be *prior*. Thus Maxwell and Boltzmann proposed the Maxwell–Bolztmann distribution several decades before it was possible to check it experimentally—and even before the hypothesis of the existence of atoms and molecules was generally accepted. A prior distribution is not something to be accepted or rejected on subjective grounds, but a hypothesis to be put to the test. Guessing prior probability distributions is methodologically akin to guessing any other functional hypotheses; and guessing probability values is no different from guessing distances or ages. All such judgments, whether of probabilities or of sizes, are guided by incomplete data and by hunches (sometimes called 'heuristic rules') and are, of course, subject to error (Tversky and Kahnneman, 1974). There is nothing wrong with such errors in subjective estimates of probabilities, or of anything else, as long as we have the chance of revising them—which is usually the case in science and technology but not so in everyday affairs, where bad mistakes of the sort are often paid dearly.

What is definitely wrong is the subjectivistic (or personalist) doctrine of probability (de Finetti, 1972; Savage, 1972). According to this doctrine probabilities (*a*) express nothing but credences, and consequently (*b*) they are always estimated subjectively. If a scientist or a technologist were to proceed in this way he would be charged with violating the ideal of objectivity. The theoretical physicist, chemist, or geneticist who calculates the probability of the transition of a system from one state to another does so not on a valid hunch but on the basis of law statements such as the Schrödinger equation. And he takes it for granted that, if the scientific theory is correct, the calculated probability quantitates the tendency or propensity of the system to make such a transition.

The science of chance was born only a few centuries ago, and it is not yet taught in elementary school. Consequently we are still unfamiliar with the

ideas of randomness, correlation, and their kin. This gives rise to the gambler's fallacy, which may lead an individual to his ruin. There is an even more basic mistake, which is betting in games of chance. Consider the so called "lottery paradox" (Kyburg, 1961), which has elicited the flow of so much philosophical ink. A person ignorant of the nature of chance bets that the (presumably fair) die he is rolling will not come up face n, where n ranges between 1 and 6. If each of these six statements is taken as true, their conjunction must be accepted as well—but of course it is false, because the 6 possible events are mutually exclusive. The paradox is dissolved on realizing that each of the above statements of the form "The die will not come up n when tossed" is groundless. A probabilistic analysis of the game of dice allows one to make only probabilistic statements, such as "In the long run this die will not come up n about five out of six tosses". The moral is obvious: Don't bet if you can avoid it.

Another example is the classical finding that most lay subjects are biased towards conjecturing associations (i.e. joint occurrences of properties or events) and neglect negative instances (Smedslund, 1963). Experienced nurses were presented with a 2×2 contingency (or presence–absence) table such as this one:

		Disease Present	*Disease* Absent
Symptom	Present	*a* 37	*b* 33
	Absent	*c* 17	*d* 13

85% of the nurses estimated that the symptom–disease correlation was strong, for they focused on the quadrant *a* with total neglect of the other three cells. The actual correlation coefficient is negligible, namely $37/70 - 17/30 = 0.038$. Ward and Jenkins (1965) found similar results working with college students presented with contingency tables concerning the association of cloud seeding with rainfall.

3.2. *Causal Hypotheses*

Whereas the statistician is happy if he finds a strong association between two variables, the scientist is unhappy if he cannot explain it, and the technologist if—for want of knowledge or of means—he cannot alter it. For example, sociologists would like to know why schooling and earnings used

to be strongly correlated. Was it that increase in schooling produces a higher earning potential or, conversely, that a higher income makes longer schooling years possible?

Statistical correlations can be explained by hypothesizing (*a*) some underlying probability distribution(s), or (*b*) a third factor influencing, or even generating, the observed associated traits, or (*c*) a causal relation between the two variables—or rather their variations. Let us deal with the latter, which are the most popular and at the same time the most reviled by the empiricist tradition, which regards the notions of cause and effect as remnants of metaphysics. (See Bunge (1959) for details on causation and its philosophy.)

The traditional causal hypotheses are of the form *A is necessary for B*. (Sufficiency may require further concurrent causes.) Since the birth of probabilistic theories in science and technology we must reckon also with semicausal probabilistic hypotheses of the basic form *The probability of A causing B equals p*. In these formulas 'A' and 'B' stand for events or kinds of events. Things and their properties do not qualify as causes and effects: only changes in properties of concrete things do. (Hence an initial state cannot be said to cause a final state—the more so since the process in question may be uncaused, i.e. spontaneous.) There would be no causation in an unchanging world, for causation is a mode of becoming, not of being.

In general, a functional relation cannot be read as a causal hypothesis without further ado. For example, although the weight of a body is a function of its volume and specific gravity, these cannot be said to cause the former: weight is an effect of the acceleration caused on bodies by a gravitational field. Some functional relations can be interpreted causally, others cannot. Take for instance a functional relation of the form "$y=f(x)$", where x and y are real variables. If x and y happen to represent events, i.e. changes of some kind, then perhaps x can be interpreted as the cause of y or conversely. If not, we may try the derived formula "$\mathrm{d}y=f'(x)\,\mathrm{d}x$", where $\mathrm{d}y$ is the variation in y corresponding to the change $\mathrm{d}x$ in x. A priori it may be that $\mathrm{d}x$ is the cause of $\mathrm{d}y$, or conversely, or that neither causes the other—i.e. that $\mathrm{d}x$ and $\mathrm{d}y$ are merely conjoined or associated. Which of the three interpretations is the correct one must be found out by additional research, either by looking for underlying mechanisms or by experimentally controlling or inhibiting changes in either variable and measuring values of the other.

Take the well-worn case of the ideal gas at constant temperature in a container equipped with a movable piston, and "obeying" the law

"$pV = \text{const.}$". If the piston compresses the gas, the decrease $-\mathrm{d}V$ in volume is the cause of the increase $\mathrm{d}p$ in internal pressure. But if the gas is allowed to push the piston, the decrease $-\mathrm{d}p$ in internal pressure will cause the increase $\mathrm{d}V$ in volume. Both causal relations are covered by the single formula "$\mathrm{d}p.V + p.\mathrm{d}V = 0$" derived from the gas law. In short, functional relations do not become causal hypotheses unless supplemented by causal interpretations. Moreover in some cases such interpretations cannot be appended. For example, Schrödinger's equation in time does not describe the causal evolution of the state function, because the latter is a probability amplitude. And even in elementary mechanics there are problems that have any desired number of solutions (Truesdell, 1974), so that classical mechanics can hardly be said to be a paragon of causal theory.

Moreover certain nonprobabilistic ("deterministic") models involving nonlinear differential equations have nonperiodic solutions that look random (and are miscalled 'chaotic'). That is, the solutions look totally irregular even to the trained statistician, although they do not contain the probability concept and, moreover, they derive from equations representing nonstochastic processes. See Figure 8.7. This is particularly the case with equations with time delays, capable of representing after-effects such as gene expression or the effect of a temporary hump in the birth curve on the market two decades later. This mock randomness is the more puzzling as only equations with time delays can adequately represent causal relations, all of which are supposed to satisfy the principle of antecedence, according to which the effect emerges after the cause. The investigation of

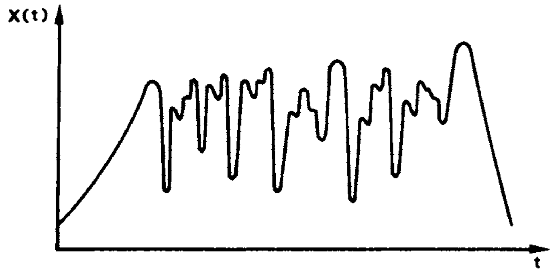


Fig. 8.7. The "deterministic" equation with time delay

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{ax(t-\tau)}{b^n + x^n(t-\tau)} - cx$$

has nonperiodic or "chaotic" solutions for large values of τ and n. See Mackey and Glass (1977). For further examples see R. H. G. Hellman (Ed.) (1980).

this problem has only begun (see, e.g. Hellman, 1980) and it is still to be exploited philosophically.

The above example shows vividly that mere empirical data cannot suggest unequivocally whether causation or mere functional relation, genuine or spurious correlation, or even chance, is at stake. A simpler example is this. Suppose you are given the sequences 0000000000 and 0100111011. You may be tempted to bet that, whereas the first was generated by some causal mechanism, the second was produced by a randomizer. Yet both sequences have the same probability, namely 2^{-10}. In sum, data are not enough to tell the difference between randomness and causation. Nor does intuition, even when accompanied by experience, suffice. We also need some hypothesis about the underlying mechanism—a testable conjecture of course—the more so since certain nonprobabilistic mechanisms give rise to behavior that looks random.

Whereas statisticians are usually content to find correlations, scientists and technologists are interested in discovering whether there is a functional relation, perhaps even a causal one, behind a regression line. This is a tricky problem that a mere contemplation of correlations cannot solve. In fact, although weak correlation disproves conclusively functional relation and causality, strong correlation may suggest either without proving them. Thus the strong correlation between per capita GNP and automobile production can be explained, *prima facie*, in either of the following alternative ways: (*a*) rising GNP causes rises in automobile sales—but then why does it not cause a similar increase in the sales of philosophy books?; (*b*) GNP and automobile sales are two different indicators of industrialization—but then why are they not similarly associated in the industrialized socialist countries?; (*c*) the growth of cities and the inadequacy of mass transportation systems cause the need for more automobiles, a need that can be satisfied by a rising per capita GNP—but then why do the Western European nations exhibit a similar correlation while having generally adequate collective transportation systems? In sum, although strong correlation may suggest causation it does not indicate its direction and, of course, it does not prove it. Monetarist economists have still to learn this lesson, for they like to buttress their hypothesis of the causal efficacy of money with some favorable correlations while disregarding the unfavorable ones. (See Modigliani (1977) for the negative correlation between monetarist policies and prosperity in the post-war U.S.A.)

But why bother with causality if quantum theory, the most fundamental physical theory, is probabilistic? Does this not prove that causality is a

myth? This was indeed the impression initially given by the quantum-theoretical revolution, particularly when allied with the positivistic criticism of causality. But on closer examination quantum physics is seen to have causal as well as probabilistic ingredients. Indeed probabilities evolve under the influence of forces—and forces are causes. (More precisely, the form of the probability amplitude depends critically upon the form of the hamiltonian, which represents the forces among the components of the system as well as the forces exerted by the environment on the system.) For example, when a physicist calculates or measures the probability that an atom will scatter an incoming electron within a given solid angle, he determines the propensity of a given event (the two entities coming close together) *causing* another event (the scattering). In short causality has blended with probability.

What holds for physics holds a fortiori for all the other factual sciences. In all of them we find causal hypotheses, sometimes including a chance ingredient, as in the schema "The probability of A causing B is p". Causality, far from having disappeared, is alive and well though now as a component of a richer determinism (Bunge, 1959a, 1982a).

4. Requirements

4.1. *Strength and Precision*

We make conjectures all the time but, unless we are fools, we do not assign them all the same status. Upon examination some hypotheses prove to be weaker, or wilder, or less deep, or less testable, or less true, or less useful than others. We should accept only those that pass certain tests—and even so we may have to demote or even discard some of them as new knowledge becomes available. Now, there are several kinds of hypothesis and therefore, presumably, different sets of requirements. Therefore we should start by distinguishing hypotheses with regard to their methodological status.

Our very first distinction is between substantive and nonsubstantive hypotheses. A *substantive* hypothesis is a hypothesis that describes, explains, or forecasts facts of some kind. The phenomenological and mechanismic hypotheses, as well as the probabilistic and causal ones, are substantive. On the other hand the *nonsubstantive* hypotheses are valuational or prescriptive: they guide research and action. We distinguish two kinds of them: methodological and axiological. A *methodological* (or

instrumental) hypothesis is a tentative statement concerning the methodological status of a statement, a procedure, or a plan—e.g. that a certain formula is, or is not, a suitable axiom or a convenient definition, or that a certain procedure is, or is not, adequate for attaining some goal, or that a certain political tactic is likely to bring victory. And an *axiological* (or *valuational*) hypothesis is a value judgment—e.g. that a given theory is all round better than another, or that a certain action or a given life style is good. Obviously, we make hypotheses of all these kinds in all walks of life, from pure mathematics to mass murder.

We should impose two requirements on hypotheses of all kinds, whether substantive or nonsubstantive. Of two rival hypotheses we should prefer the better grounded or justified, and the one belonging to the wider conceptual system. (This is a norm, not a hypothesis.) The former condition because rationality entails, among other things, the rejection of authority, even that enjoyed by success. The second condition because the wider the conceptual system the more propositions can check the given hypothesis. More on these conditions later on.

Aside from those two conditions, we treat differently hypotheses of different kinds. We prefer the strongest, more precise, deepest, more systematic, and better confirmed substantive hypotheses. These criteria are epistemic, not utilitarian. On the other hand what we expect from nonsubstantive hypotheses is effective guidance. Thus from methodological hypotheses we expect that they help us solve methodological problems of some kind, such as the organization or test of a body of propositions, the gathering of data, the design of artifacts, the outlining of plans, or what have you. And from axiological hypotheses we expect that they help us choose or do the best or avoid the worst. In the remainder of this section we shall deal only with substantive hypotheses, and shall do so in several respects.

From a logical point of view hypotheses may be ordered in respect of strength. A formula A is said to be *logically stronger* than a formula B if, and only if, A entails B. (In turn A entails B, or $A \vdash B$, only when the conditional "If A then B" is logically true, or holds for all interpretations of A and B.) Thus p is logically stronger than "p or q", where q is an arbitrary proposition, i.e. one not necessarily related to p; "p & q" is stronger than either p or q; and an "all" statement is stronger than the corresponding singular statement, which in turn is stronger than the corresponding "some" statement.

Paradoxically, the strongest of all formulas are the logically false ones, i.e. the contradictions, for they entail anything. On the other hand the

tautologies, or logically true formulas, are the weakest of all, for they entail nothing but further tautologies, and they follow from no premises. Logic, in short, is self-validating and self-contained—in fact it is the only such epistemic field. For this reason it cannot be invalidated by any factual discoveries. However, this does not entail that logic is detached from the rest of human knowledge. Far from it, logic underlies (is presupposed by) all rational discourse. In particular, every consistent theory contains logic as a proper part. More on logic in Vol. 6.

Whereas in logic we seek to establish the weakest formulas, in all other fields of research we steer a middle course between tautology and contradiction. However, with regard to logical strength there is a marked difference between the course taken in the theoretical branches of a science and that taken in its empirical (laboratory or field) ones. (Except for the formal sciences, all the others have both theoretical and empirical branches. Thus physics is composed of theoretical and experimental physics, and biology of mathematical biology and experimental biology.)

Indeed, in theoretical science we should prefer the *strongest* noncontradictory hypotheses compatible with certain general principles as well as with the relevant examples (in the case of mathematics) or empirical data (in the case of factual science). On the other hand, laboratory and field research should seek to establish the *weakest* nontautological formulas compatible with the most general principles of science, such as the principle of conservation of energy. Such are the data contained in laboratory protocols and in the notebooks of field researchers. They are in principle corrigible in the light of further data or of well-confirmed hypotheses but they do not replace or entail hypotheses.

Let us consider *precision* next. Like concepts, hypotheses come in various degrees of precision, from the ones couched in ordinary language to those couched in a mathematical language. Thus compare "There must be some thing, somewhere, that is doing something to this thing", with "Set A is included in set B". Of all the ordinary language expressions, counterfactuals break the record of imprecision. Thus the sentence 'If A were the case then B would happen' can be interpreted in a number of ways, such as "If A then B, but A is false", and as an injunction to bring about A so as to attain B. This is why counterfactuals do not occur in scientific theories and why it is wrong-headed to try and build exact theories about them.

The most precise formulas are of course those containing exact concepts. But even among mathematical formulas we can distinguish different

degrees of precision. Thus an equation is more precise than the corresponding inequality, and a quantitative formula more precise than the corresponding qualitative ones. Compare "f decreases monotonically" with "For all real $t \neq 0, f(t) = 1/t$". There are infinitely many functions that fit the former description, only one that fits the latter. Consequently the chances of being right are greater when proposing an imprecise hypothesis than when proposing a precise one. But scientists are not particularly interested in cheap truths, just as moralists are not interested in passing through wide doors. Other things being equal they prefer the more precise hypotheses because they "say" more.

The preference for precision over imprecision is justifiable only if there is an option. When there is no option we must settle for imprecision—for the time being. This is unavoidable, nay desirable, at the beginning of a research project, when we formulate *working* (or *programmatic* or *schematic*) hypotheses that guide the investigation. Thus the hypothesis that every mental phenomenon is (identical with) the specific activity of some plastic neural system, summarizes an entire research program in physiological psychology—a very fertile one indeed.

Three kinds of working or programmatic hypotheses are of particular interest: functional schemata, existential conjectures, and negative ones. A *functional schema*, such as "Y is some function of X", cries for precision and suggests the search for the exact functional relation between X and Y. It is thus part of a research project aiming at substituting a precise function (e.g. the exponential one) for the indeterminate function symbol F in "$Y = F(X)$". Or take the *existential* hypothesis "There is one genotype for every phenotype". This (partially true) working hypothesis inspires the whole of genetics: in fact, a central task of geneticists is to map gene complexes into phenotypes—a task that is unlikely to be ever completed, so many are the biospecies and their traits. Finally, consider the *negative* hypotheses that there are no microphysical inertial systems, and that the speed of light *in vacuo* fluctuates around c instead of being constant. These suspicions could generate a research line resulting in a generalization of relativistic physics. (Being negative hypotheses they can belong to no theory although they may suggest building a theory complying with them.) In sum, we cherish working hypotheses as parts of research projects. The advancement of knowledge consists, among other things, in a march from programmatic hypotheses to precise substantive ones and, in general, a progression from imprecision to precision.

4.2. *Systemicity and Scrutability*

A third property of hypotheses we must consider is their systemic status, i.e. whether or not they belong to an organized body of knowledge, such as a classification or a theory. Most of the hypotheses we formulate in daily life are stray. Thus we may conjecture that a colleague's dark glances at us are due to jealousy, although we have no theory of jealousy and no reliable jealousy indicators or objectifiers. In advanced science and technology we prefer *systemic* hypotheses, i.e. conjectures belonging to conceptual systems, and distrust the nonsystemic or *stray* ones. The advantages of systemic over stray hypotheses are obvious: (*a*) the hypotheses in a system can join to generate (logically) further hypotheses, and (*b*) systemic hypotheses are supported not only by whatever empirical data may be compatible with them, but also by the other components of the system. See Figure 8.8.

Stray hypotheses may concern (refer to) a large or a small domain of facts. The stray hypotheses that account only for a small subset of the total set of facts it addresses are called *ad hoc* (for that). Here are three examples of stray hypotheses that are also *ad hoc*. (*a*) "Criminals are the product of society". (The same society "produces" a majority of non-criminals.) (*b*) "Underground and cave dwellers have lost their sight for lack of use of their eyes". (Not all such animals are blind. Besides, genes are presumably not affected by luminosity.) (*c*) "The dinosaurs became extinct because their brains were too small to cope with a complex environment". (The dinosaurs had been remarkably successful, i.e. prolific, before the end of the Cretaceous period—at which time many other species became extinct as well.)

Whereas some *ad hoc* hypotheses are proposed to "save the appearances", i.e. to cover or even explain some group of facts, others are

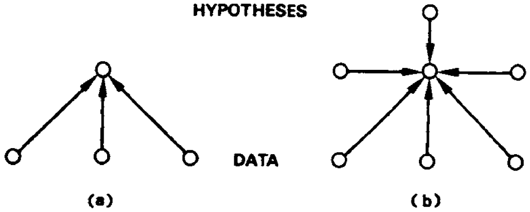


Fig. 8.8. The systemicity of hypotheses. A stray hypothesis (*a*) is supported at best by the data favorable to it. A systemic hypothesis (*b*) is supported also by the other hypotheses in the same conceptual system. The arrows indicate support, not inference.

introduced to save further hypotheses endangered by unfavorable evidence. We call the latter *protective* hypotheses. They are not all alike: whereas some may turn out to be true, others are not even scrutable. We call *bona fide* protective hypotheses those scrutable conjectures that are designed to save other hypotheses, and *mala fide* those which, being introduced for the same reason, are inscrutable. Whereas the former can in principle be shown to be true or false to some extent, the latter cannot.

The hypothesis that the naturally occurring elements are mixtures of different isotopes was initially framed to save the atomic hypothesis: it was testable and eventually found true. Likewise the hypothesis that an organism does not respond in the same way every time it is subjected to a given stimulus, for it keeps changing its internal state, was originally introduced to save stimulus–response psychology but it turned out to be true. Both are examples of *bona fide ad hoc* protective hypotheses. On the other hand the psychic's claims that his failures are due to somebody's hostility inhibiting his paranormal abilities, and the psychoanalytic fantasy that most people do not exhibit their Oedipus complex because they have repressed it, are paragons of *mala fide* protective hypotheses.

Mala fide protective hypotheses are peculiar to everyday reasoning and pseudo-science. On the other hand *bona fide* protective hypotheses are very common in science, technology, and the humanities. Sometimes they are introduced to discard data that collide with well established theories. (Imagine an engineer rejecting classical mechanics because his bridge collapsed.) At other times *bona fide* protective hypotheses are framed to shield new theories from rash criticism. (Imagine Darwin giving up his theory of evolution because of the gaps in the fossil record.) There is nothing wrong with a protective hypothesis as long as it is *bona fide*, i.e. independently testable. (For further examples see Bunge, 1967a, Ch. 5, Section 5.8, and 1973b, Ch. 2, Section 1.) We shall come back to this question in Ch. 11, Section 2.1.

The fourth and last property of hypotheses we need to look into in this chapter is scrutability; truth and utility will be examined later on. A substantive factual hypothesis can be said to be *scrutable* or *testable* if, and only if, (*a*) it refers exclusively (whether truly or falsely) to real (material) entities, and (*b*) it can be checked for truth by contrasting it with data or with the bulk of accepted knowledge. Otherwise the hypothesis is inscrutable or untestable. A hypothesis that is both inscrutable and stray (nonsystemic) may be called *wild*. Typically, pseudo-scientific speculation is wild.

Scrutability, a methodological predicate, does not coincide with intelligibility, a psychological one. A bold new hypothesis may be both scrutable and hardly intelligible to but a few initiates. Nearly every profound innovation is baffling: "When the great innovation appears, it will almost certainly be in a muddled, incomplete and confusing form. To the discoverer himself it will be only half-understood; to everybody else it will be a mystery. For any speculation which does not at first glance look crazy, there is no hope" (Dyson, 1958, p. 80). There is no great and interesting research, in any field, without speculation; the only condition we must impose on it is that it be scrutable, not wild.

Some sciences are and will probably remain largely speculative. Cosmology, palaeontology and history are among them. Take for instance the speculations of palaeontologists. Far from being wild they are guided by (*a*) fossils, (*b*) associated remains (e.g. those of plants and prey), (*c*) considerations of homology with living organisms, (*d*) geological data and hypotheses, and (*e*) general biological and even purely mechanical principles concerning, e.g., posture and locomotion. For example, it is now possible to argue for the hypothesis that all of the great dinosaurs were water dwellers, as well as for and against the fantastic (but not wild) conjecture that dinosaurs, unlike their modern successors, were warm-blooded and therefore had a quick metabolism. Thus it is now possible to cast "a cold look at warm-blooded dinosaurs"—the title of a recent scientific symposium (Thomas and Olson, 1980).

To conclude. When considering hypotheses for test or adoption we should generally prefer depth to superficiality, logical strength to weakness, precision to inexactness, systemicity to isolation, and scrutability to untestability. However, imprecise conjectures are acceptable if they are promising research guides, i.e. programmatic hypotheses. And superficial and relatively weak hypotheses may be unavoidable if we need quick practical results. What we cannot accept under any circumstance, in any walk of life, are inscrutable hypotheses, particularly of the protective kind, for they are the mark of dogmatism.

5. Concluding remarks

Dogmatic rationalists, intuitionists and radical empiricists have no use for hypotheses: the former own indubitable a priori axioms, the intuitionists are endowed with a special direct insight (some of them even with what Husserl

called *Wesensschau*, i.e. vision of essences), and the radical empiricists, more modest, are content with perceptions and images. Even the great Russell, during his logical atomism period, proposed to eliminate what he called 'inferred entities' (i.e. hypothesized objects) in favor of constructions out of certified, in particular perceptible, entities. He proposed the rule "Whenever possible substitute constructions out of known entities for inferences to unknown entities" (Russell, 1914). And he applied this maxim to the analysis of matter, which he dissolved into phenomena. In a similar vein, operationists and members of the Vienna Circle attempted to characterize electrons and other transempirical entities, not in terms of their objective properties—the way the theories of matter do—but in terms of pointer readings or some other empirical operations. Fortunately scientists did not follow up this anthropocentric program—except of course during their occasional forays into philosophy. They went on framing and checking increasingly sophisticated hypotheses about hypothesized (but supposedly real) entities such as fields and subatomic particles, extinct organisms and societies, the interior of stars and of brains, and so on. Humanists and technologists behave similarly: far from avoiding conjectures they frame them all the time; they try to avoid only wild hypotheses.

Hypothesizing is just as necessary as gathering data, because to investigate is to handle problems, and to solve a problem is to frame a hypothesis, or find a set of data, that may satisfy the condition (generator) of the problem. Hypotheses fulfill a number of functions in ordinary life, science, technology, and the humanities. Some of them summarize experience and others generalize it. Some allow us to deal with real or presumptively real but imperceptible facts, still others trigger reasonings or fill gaps in them. Finally, others guide (or block) entire research plans, in particular the design of observations. The services hypotheses render are so many and important that any attempt to dispense with them would be as silly and ineffectual as trying to do without any data.

That the ultimate components of the universe are fields and particles, and all of these are restless; that the sun and the planets are roughly spherical and compose a system; that cells are the units of life and all modern organisms descend from others that lived long ago; that every society is composed of an economy, a polity and a culture—however primitive—and that these subsystems interact with one another vigorously—all these are hypotheses, not data. In fact most of our scientific

knowledge is hypothetical, though of course not all conjectures are equally tentative: some of them—such as the ones we have just listed—are as well established as the firmest of data. They are supported not only by data but also by further hypotheses belonging to some conceptual system or other. Conceptual systems is what we shall study next.

CHAPTER 9

SYSTEMATIZING

All humans crave for knowledge, and not just any old heap of bits of information but systems, i.e. sets of epistemic items possessing unity and coherence. The easiest way to obtain both unity and coherence is through radical simplification and reduction to a few basic ideas. This is the way nonscientific ideologies are fashioned.

Simplification is unavoidable and reduction is desirable—up to a certain point, beyond which they involve inadmissible distortions. Reality is much too varied and changeable, creativity much too incessant, and rationality much too demanding, to tolerate simplistic systems. We need conceptual systems capable of accounting for diversity and at the same time possessing the desirable properties of unity and consistency. These systems are classifications and theories.

Despite the craving for conceptual systemicity, there is still some distrust for classification and theorization. We often say we do not like being pigeonholed, and that a gram of practice is worth a ton of theory. This distrust has several sources: sheer anti-intellectualism, misunderstanding of the nature of classification and theory, and ignorance of their role in civilized life. In this chapter we shall study classification and theory, without which man would lead a rather brutish existence, for they help us understand the world and transform it.

We distinguish different degrees of epistemic systemicity or cohesion: the mere collection of bits of information, the context, the classification, the doctrine, the theory, and the system of theories. A catalogue, such as a telephone directory, is an instance of an unstructured set of epistemic items. A context is a more or less loosely knit collection of propositions referring to a single domain. Stories, flow diagrams, wiring diagrams and organigrams are instances of contexts. A context need have no logical structure, so that knowledge of one item of it does not allow one to infer anything about other members. But, by virtue of its definite reference, a context has referential or ontological unity. In Vol. 1, Ch. 2, Section 3.4, we characterized a conceptual context as an ordered triple $\langle P, Q, R \rangle$, where P is a set of propositions, Q the set of all predicates occurring in P, and R the set of referents of the predicates in Q.

Classifications, doctrines, theories (in particular models) and systems of theories are contexts possessing logical as well as referential unity. A classification is a system of propositions concerning items of some kind (e.g. organisms) and their inclusion relations (e.g. "Species S is included in genus G"). A doctrine is a context such that some of its propositions are logically related to others: it need not constitute a hypothetico-deductive system and it need not even be consistent. A theory is a hypothetico-deductive system, i.e. a system of propositions every one of which is either a premise or a logical consequence of a group of premises. And a system of theories is a collection of mutually consistent theories and of bridges among them.

1. Classification and Theory

1.1. *Classification*

It seems that all sentient organisms are capable of perceiving certain types and patterns—visual, auditory, tactual, gustatory, olfactory, spatial, temporal, and others. Now, a *type* is whatever is shared by a mass of apparently unrelated items, and a *pattern* is whatever is invariant in a mass of items. Thus all human faces, however great their individual differences, have the same type or general shape; and all waves fit the same general pattern even though they differ in amplitude and frequency. Since we can perceive types and patterns, by the same token we can discard individual differences and thus throw some information away. We do this either from the start or as a result of analysis. In the former case we grasp types and patterns in a synthetic way, sometimes intuitively; in the latter case we conjecture, test, and conjecture again until we grasp type or pattern. In this section we shall study typification, i.e. the formation of types and systems of types.

We saw in Ch. 5, Section 1.2 how classes are formed, namely by grouping together individuals that share certain properties, such as having the same mass or the same descent, even if they differ in all other respects. A single attribute A and its complement not-A allow one to make black or white statements such as "c is an A" and "c is a non-A". A pair of attributes, A and B, allow us to form four different propositions: "c is an A and a B", "c is an A and a non-B", "c is a non-A and a B", and "c is a non-A and a non-B"—i.e. the components of a 2×2 contingency table. In general, for n attributes we may construct 2^n propositions for any given individual. Each

such set is called a *Boolean partition*. We use such qualitative partitions every day.

For certain purposes we must use quantitative predicates, such as distances and probabilities, which allow us to form infinitely many propositions, and so infinitely fine partitions. Think, e.g. of the infinitely many distances a moving point can be from a fixed center. We call these *Cartesian partitions*. See Figure 9.1. Because a Cartesian partition contains infinitely many cells, it provides a far richer representation of reality than any Boolean partition—so rich, in fact, that it incorporates information that cannot be obtained by observation and measurement. (Any run of measurements yields only a finite number of numerical values.) It is thus a formidable analytic tool and, by the same token, an inadequate categorization tool, for categorizing is lumping as well as splitting. Let us take a closer look at classification, whose principles are often misunderstood by its most conspicuous users, namely the biologists.

A *classification* of a given collection of individuals—be they concepts, concrete things, events, or what have you—is a conceptual operation with the following characteristics:

(i) each member of the original collection is assigned to some *class*;

(ii) each class is composed by some of the *members* of the original collection, and no class is composed of subclasses;

(iii) each class is a *set* the membership of which is determined by a predicate or a conjunction of predicates;

(iv) each class is *definite*, i.e. there are no borderline cases—which is ensured by employing only definite or exact predicates and avoiding vague ones such as "young" and "distant";

(v) any two classes are either mutually *disjoint* (i.e. have no members in common) or one of them is *included* (contained) in the other: if the former

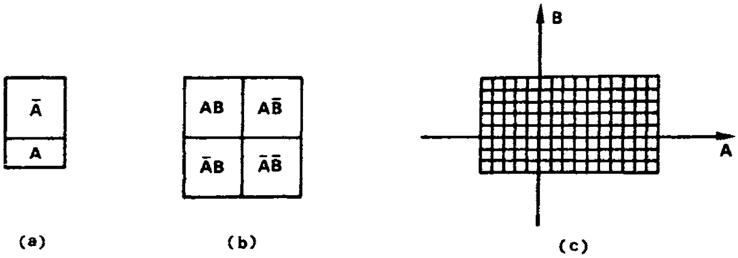


Fig. 9.1. Categorizations. (*a*) Dichotomy. (*b*) A Boolean partition on the basis of two qualitative predicates. (*c*) A Cartesian partition on the basis of two quantitative predicates.

they are said to belong to the same *rank*, otherwise to different ranks;

(vi) only two relations are involved: the *membership* relation ($\in$) holding between the individuals of the original collection and the first rank classes, and the inclusion relation ($\subseteq$) that relates classes of different ranks (e.g. species to genera);

(vii) every class of a rank higher than the first equals the *union* of some or all classes of the immediately preceding rank (e.g. every genus equals the union of its species);

(viii) all the classes of a given rank are *pairwise disjoint* (do not intersect), so that no item in the original collection belongs to more than one class of the same rank;

(ix) every partition of a given rank (e.g. every grouping of the original collection into species) is *exhaustive*, i.e. the union of all the classes in a given rank equals the original collection;

(x) a classification violating any of the above conditions must be repaired or given up.

Clearly dichotomies, such as existent/nonexistent, satisfy all of the above conditions. But they are trivial classifications because they involve a single rank, whereas nontrivial classifications are "vertical" as well as "horizontal". This requires seizing on more than one property. (For example, two attributes, P and Q, elicit the formation of two species, namely $S_1 = \{x|Px\}$ and $S_2 = \{y|Qy\}$, and one genus, namely $G = \{z|Pz \text{ or } Qz\} = S_1 \cup S_2$.) Now, unless the two predicates are mutually incompatible, such as sitting down and running, they may induce the formation of partially overlapping classes. Therefore seizing on arbitrary predicates will not produce a classification proper except by accident.

The only method for producing a classification proper, i.e. a categorization satisfying the nine conditions listed above, is the one based on *equivalence relations* such as those of congruence, homology, common descent, and same age bracket. In fact every such relation splits the set on which it is defined into mutually disjoint (exclusive) and exhaustive (inclusive) sets. These sets are called *equivalence* (or *homogeneous*) classes, and the family of such classes the *partition* of the given set S by the given equivalence relation $\sim$, or $P = S/\sim$. See Figure 9.2a. For example, T shirts for men come in three different sizes, so that the relation "the same size as" ($\sim$) partitions T thus: $\mathrm{T}/\sim = \{\mathrm{S},\mathrm{M},\mathrm{L}\}$, where $\mathrm{S} \cap \mathrm{M} = \varnothing$, $\mathrm{S} \cap \mathrm{L} = \varnothing$, $\mathrm{M} \cap \mathrm{L} = \varnothing$ (condition viii above), and $\mathrm{T} = \mathrm{S} \cup \mathrm{M} \cup \mathrm{L}$ (condition ix). In this case the species are S, M and L, and the genus coincides with the original set T. A more sophisticated example is the formation of clades on the basis of

homologies such as that between the forelimbs of man, seal, and bat.

Obviously there are as many partitions as equivalence relations. Thus when classing fossil vertebrates one may focus on the conformation of the hip bones, or the possession or lack of a predentary bone in the lower jaw, or the conjectured mode of locomotion, or the conjectured habitat, etc. Each equivalence relation induces its own partition of the given set. Consequently using n different equivalence relations $\sim_1, \sim_2, \ldots \sim_n$, we obtain n partitions $P_i = S/\sim_i$, some of which may be identical. This family of partitions may be pictured as a stack of disks, every one representing the same original pie cut in different ways. See Figure 9.2b.

Once a first partition $P_1 = S/\sim_1$ of the original collection has been effected, we have completed the first rank of the classification: it consists of

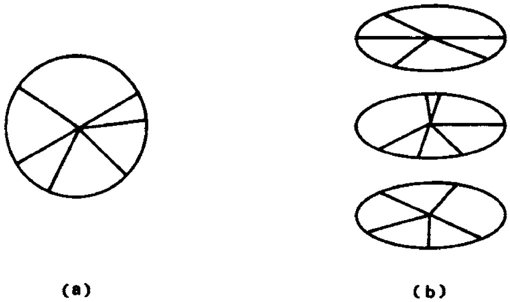


Fig. 9.2. (a) Partition of a collection by an equivalence relation. Each sector (subset) is homogeneous in the given respect, and the union of all the sectors equals the original set. (b) A family of partitions of a given set.

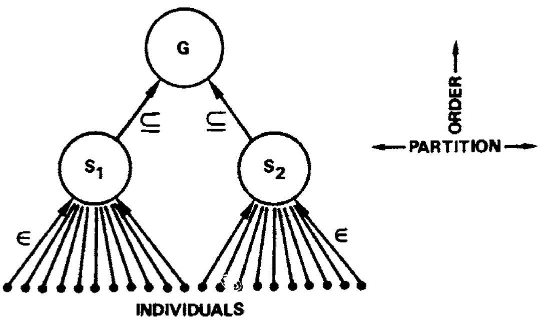


Fig. 9.3. Individuals, species, and genera. Each individual belongs to ($\in$) some species (lowest order taxon), and each species is included ($\subseteq$) in a genus.

species (or subspecies, or varieties). These species may be grouped in turn by using a second equivalence relation $\sim_2$, this one defined on the family P_1, i.e. $P_2 = P_1/\sim_2 = \{G_1, G_2, \ldots G_m\}$. The members of this partition are called *genera*. A third equivalence relation, this one defined on P_2, will induce the formation of the third rank of the classification, namely that composed by *families*—and so on. See Figure 9.3.

To recapitulate, these are the steps in a classification process:

0. A set S of individuals is given or produced.
1. Find a "significant" equivalence relation $\sim_1$ on S.
2. Partition S by $\sim_1$, i.e. form the collection of species $P_1 = S/\sim_1 = \{S_1, S_2, \ldots S_n\}$.
3. Find a "significant" equivalence relation $\sim_2$ on P_1.
4. Partition P_1 by $\sim_2$, i.e. form the family of genera $P_2 = P_1/\sim_2 = \{G_1, G_2, \ldots G_m\}$, such that every member of it be the union of some (one, two or more) species.
5. Repeat the process for the higher ranks as far as necessary or convenient.

One may also invert the process, i.e. start by partitioning the original collection into genera, or even higher rank classes, and proceed to distinguishing subsets within them. For example, the fisherman probably starts by partitioning all fish into edible and inedible—an artificial partition for it is anthropocentric not ichthyocentric. The equivalence relation he uses is "as edible as". He may then proceed to divide the class of edible fish into those which are easy to fish and those which are not. This second equivalence relation, namely "as easy to fish as", is said to *refine* the former. To take a more interesting example, suppose we start with the order primates, and distinguish within it the family of hominids, which in turn includes the genus *Homo*, which in turn includes our own species *Homo sapiens sapiens*. Implicit in this classification are three different equivalence relations that are not always spelled out explicitly. (In fact most biologists partition and order in a rather intuitive way, not caring much for taxonomy, or the methodology of classification. They often pay for this with mistakes or barren controversies. See, e.g. Ghiselin (1981) and the ensuing discussion.) As we descend from the higher to the lower rank classes we do so with the help of finer and finer equivalence relations.

Every classification has two dimensions: the horizontal one attached to the relations of membership ($\in$) and equivalence ($\sim$), and the vertical one linked to the inclusion relation ($\subseteq$). Every good classification is then a

trivial example or model of elementary set theory. But not all classifications satisfy every one of the nine conditions we listed a while ago. In particular, some classifications are made with the help of vague or inexact predicates, such as color and shape predicates. And others, actually the great majority, are not exhaustive for want of information. (We keep discovering, inventing or making objects belonging to formerly unknown species.)

Even if a partition is performed with the help of an exact predicate, and therefore an exact equivalence relation, the result may be unsatisfactory because the predicate may be superficial and therefore incapable of bringing about the unity or systematicity we want. For example, the original nomenclature of biocatalyzers (enzymes) was based on the equivalence relation "acting on the same molecule". I.e., if X names the molecule on which enzyme Y acts, then the name of Y is X-*ase*. (Thus arginase catalyzes the hydrolysis of arginine to ornithine and urea.) As the number of known enzymes climbed up to the thousands this nomenclature became impractical. The International Enzyme Commission produced a systematic partition, hence nomenclature, based on a functional description of enzymes. In other words, the new equivalence relation is "having the same function". For example, all the enzymes that catalyze hydrolises are called *hydrolases*, and those that isomerize, *isomerases*.

A clear case of superficiality is that of *artificial* partitions, i.e. those based on traits that, far from being primary and intrinsic, are secondary or concern our relation to the objects concerned (e.g. their external appearance or usefulness). Inevitably, empiricism and pragmatism enjoin us to stick to artificial classifications. And nominalists do not care what classification one may adopt for, according to nominalism, a general sign, such as 'human', does not designate a natural kind but denotes severally the individuals it points to. In other words, according to nominalism the *fundamentum divisionis* of every classification is arbitrary, without a real counterpart: every classification is artificial, and so no classification is better than any other one.

On the other hand, realism recommends natural classifications and, among these, prefers those based on deep or essential properties, in particular laws. (For the notion of a natural kind see Vol. 3, Ch. 3, Section 3.3, and for a defense of nomological essentialism, *ibid.* Ch. 2, Section 4.2.) Essential properties are the basis of essential equivalence relations, which in turn elicit essential or basic partitions—as opposed to accidental or superficial ones. In short, classifications come in several depths, and we should prefer the deepest of all for being the more realistic. (For the notion

of depth recall Ch. 8, Section 2.2.) This does not entail that the classes themselves can be more or less real: every class is a concept, but whereas some concepts represent reality others do not. Thus, whereas the set formed by this book and the reader's last dream does not constitute a natural kind, all the bits of copper, and all the hominids and their descendants, do.

Now, mere perception is bound to lead to superficial partitions. Deep partitions call for hypotheses and, in particular, law statements, i.e. formulas about patterns. And, because law statements belong (by definition) to theories, if we want deep classifications we need theories, the deeper the better. Good examples of the power of theory to inspire deep classifications are contemporary biological systematics (based on the theory of evolution), the periodic table of the elements (based on the atomic theory), the classification of hadrons based on the quark model, and the classification of materials based on their constitutive relations or specific laws.

Classing and theorizing are then mutually complementary activities. Categorizing precedes theorizing if only because every theory is about some category of objects. In turn, theory allows one to refine the coarse and shallow pretheoretical classifications. Moreover a classification *is* a theory of a kind. Indeed, unlike the propositions in a catalogue, those in a classification are logically related: they compose a system with logical unity. Thus from the statements that all mammals are vertebrates, and all vertebrates animals, it follows that all mammals are animals. To be sure this piece of knowledge is not a surprising new theorem but was available from the start, but the point is that it is a systemic proposition, not a stray one. Into a classification we pour all we know about certain kinds, and from it we expect no new knowledge. Classifications summarize and order available knowledge, and thus comply with the requisite of empiricism, which prohibits speculation.

Classifications may be dubbed *finite theories*, in contrast to ordinary theories, which contain infinitely many propositions, so that by investigating them we are bound to obtain propositions we did not know before. (See the next section.) For the same reason classifications have a very limited explanatory power. For example, the statement that men and monkeys are primates explains our similitude to our cousins, but only because this similarity was used to construct the order of primates. And what predictive power a classification may have is strictly heuristic, for it stems only from the gaps (e.g. missing links) in it rather than from any explicit statements. For example, the theoretical classification of hadrons based on the quark

model predicted the existence of the omega-minus particle, which was eventually found in 1964. A complete classification of known items has no predictive power at all. If we want prediction and explanation we must progress from classification to theory. This we shall do anon.

1.2. *Theory*

From a logical point of view a theory is a logically organized set of statements concerning objects of some kind, i.e. it is a systematic context. At the beginning of this chapter we recalled that a context is defined as an ordered triple of sets: statements-nonlogical predicates occurring in the latter-domain (reference class) of such predicates, or $\mathscr{C} = \langle P, Q, R \rangle$ for short. Since a theory is a logically organized context, it can be characterized as a quadruple $\mathscr{T} = \langle P, Q, R, \vdash \rangle$, where $\vdash$ is the entailment relation that glues the members of P into a system. More precisely, a theory is a *context closed under deduction*, i.e. such that every statement in it is either a premise or a deductive consequence of a set of premises.

If the theory concerns exclusively conceptual objects, such as numbers or philosophical ideas, it is said to be *formal*. If the domain or reference class R of a theory contains factual items, such as molecules or societies, it is said to be *factual*. (Whereas a formal theory contains no reference to factual items, a factual theory may contain not only factual statements but also statements about properties of its own concepts. Thus a theory of change will contain mathematical statements concerning the properties of the functions representing changing properties.) The root difference between a formal and a factual theory lies in the domain R of individuals, or reference class. This difference is semantic and it will occupy us in Section 2.2. In the present section we shall focus on the structure of theories.

That a theory is a logically organized context means that (*a*) a theory is "based on" (presupposes or contains) some theory of logic, usually the predicate calculus with identity, and (*b*) whereas some formulas of the theory are premises, others are logical consequences of some such premises, or even consequences of such consequences—all of them derived with the help of the underlying or presupposed logic. The premises of a theory are either postulates (axioms) or definitions. The consequences of the postulates or definitions are called 'theorems', and the special cases or immediate consequences of axioms, definitions or theorems, are called 'corollaries'. The postulates and their consequences "say" something about the referents of the theory (or members of its domain R), or some

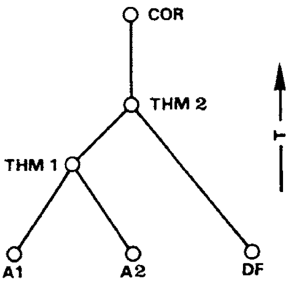


Fig. 9.4. Finite fragment of a theory based on two axioms and one definition. The total theory cannot be pictured because it is a graph with infinitely many nodes (formulas) and edges (logical relations).

concepts of the theory (members of the set Q of predicates), or about relations between concepts and referents. On the other hand a definition links two or more concepts of the theory. See Figure 9.4.

Example 1. Theory of partial order

*A*0	The predicate calculus with identity.
*A*1	R is a nonempty set.
*A*2	For all x in $R, x \leq x$ (reflexivity).
*A*3	For all x and y in $R, x \leq y \,\&\, y \leq x \Rightarrow x = y$ (antisymmetry).
*A*4	For all x, y and z in $R, x \leq y \,\&\, y \leq z \Rightarrow x \leq z$ (transitivity).
*Df*1	R is *partially ordered* $=$ R satisfies the above axioms.
*Df*2	R is a *chain* $=$ $\leq$ holds between any two members of R.
*Df*3	For all x and y in $R, x < y = x \leq y \,\&\, x \neq y$.
*Thm*1	For all x, y and z in $R, x \leq z \,\&\, z \leq y \,\&\, x = y \Rightarrow x = z$.
*Thm*2	For all x in $R, \neg(x < x)$.
Cor	For all x, y and z in $R, x < y \,\&\, y < z \Rightarrow x < z$.

This is a formal theory, although it can be interpreted in factual terms, e.g. by letting R be the set of bodies and $\leq$ the relation of being less heavy than, or R the family of nations and $\leq$ the relation of being less developed than. It is so simple and general a theory that it underlies (is presupposed by) a host of richer theories, particularly in abstract algebra. By introducing further concepts, such as those of immediate successor, first and last element, greatest lower bound, and least upper bound, richer abstract theories can be built, such as lattice theory and filter theory.

The zeroth axiom *A*0 constitutes the *formal background* of the theory, which in this case is nothing more than ordinary predicate logic with

identity. That and the other four axioms and the three definitions compose the *foundations* of the theory. In this case all the axioms concern members of the reference class and their mutual relations. The theory does not contain any semantic axioms (meaning assumptions), i.e. hypotheses about the relation between the specific predicate of the theory, namely $\leq$, and its referents or members of R. In other words, the theory is strictly formal. Its interpretation in factual terms would require the addition of two semantic postulates, one for R, the other for $\leq$—e.g. that R is a set of organisms and $<$ the relation of descent.

Example 2. Theory of exponential growth

*A*0	Ordinary predicate logic with identity plus elementary infinitesimal calculus.
*A*1	R is a nonempty set.
*A*2	F is a real valued function on the cartesian product of R by T, where T is a subset of the real line.
*A*3	F is differentiable with respect to its second argument.
*A*4	The rate F of change of F with respect to its second argument is proportional to F itself, where the proportionality constant is a positive real number depending on the individual (i.e. it is a real valued function on R).
*A*5	'T' represents time.
*A*6	'F' represents the intensity of property P of things of kind R.
Df	The relative rate of change of F equals $\dot{F}/F$.
*Thm*1	The relative rate of change of F is constant. (By A4 and *Df*.)
*Thm*2	F grows exponentially. (By *A*4 and certain principles of the infinitesimal calculus.)

The formal background of this theory is far richer than that of the previous theory for, in addition to logic, it presupposes a bulky chapter of mathematics. Thus, because F is assumed to be a continuous function, it has infinitely many values, one for each time and each value of the relative rate of change. That is, *Thm*2 is actually a nondenumerable infinity of formulas of the form "$F(t) = F(0) \exp(kt)$", where $F(0)$ is the initial value of F, k the relative rate of change, and t is in T. Contrast this infinite richness with a singular qualitative hypothesis such as "All ravens are black". (And yet our theory is about the simplest that can be stated with the help of the infinitesimal calculus.) Nor is a hypothesis of continuity necessary to build theories with infinitely many formulas. The underlying logic itself contains

infinitely many tautologies. And any theory that presupposes number theory has not only infinitely many formulas but also infinitely many concepts: ... -2, -1, 0, 1, 2, ... Such infinities are usually manageable thanks to the concept of a set and a suitable notation. Thus instead of listing all the integers, which would be impossible anyway, we introduce the single symbol '$\mathbb{Z}$'. Likewise in defining, say, the infinitely many coefficients of a power series one just writes the nth coefficient c_n.

Our growth theory is a factual theory, as shown by the postulates $A5$ and $A6$, which are semantic assumptions and, more particularly, hypotheses concerning relations between concepts in the theory and things and processes outside it. (Actually $A6$ is only a semantic axiom schema, for it specifies neither the referents nor the property P.) In other words, when joined to $A5$ and $A6$, $A4$ is a factual hypothesis and moreover a testable one. To be sure, it cannot be tested directly but only through its consequences *Thm*1 and *Thm*2. On the other hand the axioms of the theory of partial order are not testable: they may be regarded as constituting an axiomatic (hence implicit) definition of the relation "$\leq$". What can be checked is only the conjecture that this or that set satisfies those axioms, i.e. is a model of the theory of partial order. In sum, the difference in reference entails a difference in testability.

Note that our theory of growth is kinematical, not dynamical or mechanismic. Indeed it assumes no growth mechanism. On the other hand, a mechanismic theory of growth would contain statements concerning the mode of growth—e.g. by self-assembly prompted by intermolecular forces, or by cell division, or by sexual reproduction. In such a theory the law statement $A4$ would occur as a theorem, not as a postulate—and possibly in a wider form making room for random fluctuations around the central exponential trend. The demotion of axioms to theorems is a good indicator of increasing depth and therefore power. Thus the famous competitive exclusion principle of ecology, namely "Different species occupy different niches" (or "No two species occupy the same niche"), can be derived from the assumption that the members of different species have different enzymes and, as a consequence, different resource needs and different possibilities of exploiting the environment.

We have just made tacit use of the notion of subtheory, or part of a theory that is itself a theory rather than an arbitrary fragment of it. Thus a mechanismic theory of exponential growth includes, as a subtheory, the kinematical theory of growth. (In general T_1 is a *subtheory* of T_2 if, and only if, T_1 is a theory and T_2 entails T_1 or, equivalently, every formula of T_1

is included in T_2.) The subtheory may or may not have a smaller reference class than the theory of which it is a part: what matters is that all the statements of the former belong in the larger theory but not conversely. Thus particle mechanics is a (tiny) subtheory of body (continuum) mechanics, and the theory of the linear oscillator is in turn a subtheory of particle mechanics.

Subtheories should not be mistaken for *specific theories* or *models* of things of a certain kind. A model contains assumptions that do not occur in the general theory, so it cannot be part of the latter. For example, a specific theory (or model) of a fluid of a certain kind contains not only the general assumptions of mechanics but also the "constitutive relations" or special laws characterizing that fluid. Again, a theory of the evolution of plants is more specific than the general theory of evolution (on which the former is based), for it contains particular assumptions about the roles of chloroplasts and chlorophyll, roots and seeds, etc. So, it is not a subtheory of the theory of evolution. Likewise a theory of sexual selection is not a subtheory of the theory of natural selection, for it contains concepts—such as those of breeding time, competition for mating partner, and courting—that are absent from the general theory. See Figure 9.5.

In general, we may say that a theory T_1 *specializes* to theory T_2, or that T_2 is an *application* of T_1, if and only if there exists a set A of formulas such that the union of T_1 and A entails T_2. The additional assumptions may be further hypotheses (in particular indicators) or data. (Indicators or "operational definitions" are really hypotheses; they may but need not be included in the specializations or applications of a theory but they must occur in every test of it.) For example, we specialize or apply the theory of partial order when enriching it with (semantic) assumptions concerning the nature of R and $\leq$. And we apply particle mechanics when we specify the system to be a pendulum. In all specializations or applications the original reference class shrinks. That is, the special theories embrace logically the

Fig. 9.5. (*a*) Theory T_1 is a subtheory of theory T_2. (*b*) Specific theory T_2 is based on theory T_1 and contains set A of assumptions not included in T_1.

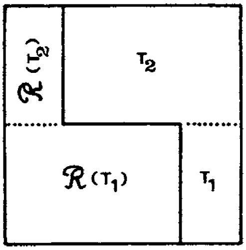


Fig. 9.6. T_2 is a specialization of T_1, and so contains more formulas than T_1 but has a smaller reference class than T_1.

general ones but are referentially included in the latter. I.e. if T_2 is a specialization of application of T_1, then R_2 is included in R_1. See Figure 9.6.

The special theories or models resulting from specialization (or application) can be called *bound models*, for they are attached to general theories. But not all special theories are based in general ones. We call *free models* those that do not result from augmenting the set of assumptions of a theory. (In other words, the totality of formulas of a bound model equals the set of all the consequences of the union of the postulates G of a general theory with the special assumptions A compatible with G, i.e. $P = Cn(G \cup A)$. On the other hand the totality of formulas of a free model is $P = Cn(A)$, for here $G = \varnothing$.)

Whereas in the advanced sciences and technologies most models are bound, in the less advanced ones most of them are free, i.e. they are built from scratch instead of on the basis of general theories. This is particularly the case with the social sciences, where there are no recognized general and valid theories. Needless to say, it takes more imagination to build a free model than to apply a general theory, for in the latter case the existing theory can be used as a scaffolding. Thus no matter how difficult it may be to build a theory of the origin of viri or of fungi, or a theory of the lineage of mammals or of the geographical distribution of plants, the contemporary theoretical biologist can avail himself of general cytological and physiological principles as well as of general theories such as genetics and the theory of evolution. On the other hand the sociologist and the economist have no such guidance. Worse, they are sometimes misguided by theories concerning ghostly entities, such as the free competitive market, or even by ideologies.

The difference between general and specific theories is not always real-

ized or understood. For example, much of the recent work on the foundations and philosophy of classical mechanics takes this to be identical with particle mechanics or reducible to the latter (e.g. Sneed, 1979; Stegmüller, 1976). It ignores continuum mechanics and in particular its constitutive relations (the law statements that characterize kinds of materials). Consequently such studies do not help us understand mechanics or, a fortiori, scientific theories in general. (See Truesdell (1983) for a devastating criticism of such pseudomechanics.) Others criticize the theory of evolution for not making precise predictions, not realizing that extremely general theories, whether in biology or in physics, are not supposed to make predictions unless considerably enriched with special assumptions and data. Thus if we wish to understand, retrodict and predict the evolution of giraffes, we must augument the general theory of evolution with a myriad hypotheses and data concerning the camelids and their presumptive habitats. That is, we must build a (bound) model of giraffes. (For an extremely simple but precise model of evolution see Appendix 1.)

Whether free or bound, the models built in factual science and technology are radically different from those studied by model theory, a branch of logic. (See, e.g. Bell and Slomson, 1969.) The latter are not models of concrete things, such as atomic nuclei and factories, but examples of abstract theories. Thus the natural numbers, together with addition (or multiplication), constitute a model of a semigroup, i.e. they exemplify the concept of a semigroup. Likewise the propositional calculus is a model of Boolean algebra, i.e. propositions satisfy and therefore exemplify the laws of Boolean algebra. This, the *logical* concept of a model, must be sharply distinguished from the *epistemological* concept of a model as a conceptual representation of concrete things of a narrow kind. But, whether logical or epistemological, all such models are conceptual: real things are not models of anything, but instead the objects of modeling. (The statement that the real world is a model of some theory makes sense only in a Platonic ontology.) This terminological remark had to be made in view of the widespread confusion between the two kinds of model introduced by the so-called set-theoretic semantics of science (e.g. Sneed, 1979; Stegmüller, 1976).

We have examined two kinds of inter-theory relation, those of subtheory to theory (e.g. statics to dynamics), and of general theory to specific theory (e.g. group theory to commutative group theory). There are several other inter-theory relations. Let us mention a few of them. *Equivalence*: Two theories have the same composition (set of formulas) but different

structures, i.e. they are organized differently—e.g. an undefined notion in one of them is defined in the other, or an axiom in the first is proved in the second. *Example*: the different presentations of the propositional calculus. *Refinement*: One theory constitutes a revision of another—e.g. it adopts a more powerful mathematical formalism, drops certain assumptions and adds others, simplifies some of them and complicates others, etc. *Example*: contemporary classical mechanics *vis-à-vis* classical mechanics. *Reduction*: One theory explains or reduces another—e.g. a black box theory is reduced to, or explained by, a mechanismic one. *Example*: molecular biology explains classical genetics. (We shall study reduction in Ch. 10, Section 3.1.) *Rivalry*: Two theories concerning roughly the same domain are inequivalent because, even though they share some predicates, they relate them differently, hence they explain or predict things differently. *Example*: the information-processing (cognitivist) and the psychobiological theories of vision.

So much for theory structure. We have presented a view that is an elaboration of the standard view adopted by mathematicians and theoretical scientists. This standard view is not accepted by all philosophers. Thus some of them mistake theories for languages, even though a language has no deductive structure and should serve to express mutually incompatible theories; others confuse theories (hypothetico-deductive systems) with hypotheses (single propositions); others try their hand at the alchemical art of dispensing with theoretical concepts in favor of empirical ones; still others, while calling themselves realists, claim that theories are metaphors; finally, others include the applications or specializations of a theory in the theory itself. (See Suppe (Ed.) (1974) for a representative sample of this population.) We shall waste no time with these quaint views for the same reason that we do not take the Flat Earth Society seriously. Those are harmless philosophical myths.

On the other hand we must sound the alarm about a real peril for all theorizing, namely the current computer cult. In fact the overrating of the computer is having extremely harmful effects on mathematics, science and technology alongside its well-known beneficial ones. One is that the general public and the bureaucracy of science feel all the more justified in their contempt for theory as their admiration for The Computer grows. This valuational mistake stems in part from the confusion between theorizing and calculation, and the enormous prestige attained by the latter in connection with space travel and missiles. It is usually overlooked that computation, whether by hand or by computer, involves both theory (e.g.

mechanics) and algorithm (e.g. the infinitesimal calculus), so no computer can replace either. Moreover, when used as calculation auxiliaries computers deliver sheer numbers that make no sense except in the light of some theory. When no theory intervenes in the input or the output of a computer, the operation has no scientific significance. Sometimes it is just a piece of pseudoscience.

A second harmful effect of the computer cult is the growing tendency among investigators to neglect the analysis, improvement, and replacement of existing theories. Indeed, a scientist devoted to computation *uses known theories* instead of examining them or inventing new ones. Not being creative, the computer cannot fill in for the original theorist; and not being critical, it cannot replace the foundations researcher. Because of this emphasis on the application of known theories, what began as the best friend of the theorist is turning against him, and what began as the computer revolution is turning into the computer counter-revolution. (More in Truesdell (1981).)

Thus far we have been concerned only with the structure and content of theories, i.e. we have dealt with the logical and the semantical concepts of a theory. But theories can be viewed in alternative ways as well, in particular methodologically, psychologically, sociologically, and historically. Thus we have at least six different concepts of any factual theory T:

Logical:	T is a set of statements,
Sematical:	T represents its referents,
Methodological:	T helps explain, predict, or design experiments,
Psychological:	T is a collection of possible thoughts,
Sociological:	T is a product of the social process of inquiry,
Historical:	T is a link in a historical process.

Different specialists are likely to emphasize different concepts of theory, and some of them are bound to claim that theirs is the only valid one. In particular, those allergic to mathematics are likely to dismiss the logical concept, and those insensitive to history will ignore the historical concept. However, the six concepts are mutually compatible, not exclusive; moreover, they complement one another. Thus the historian needs to know that a theory is (among other things) a logically organized set of statements rather than an unorganized one, let alone a single statement, for otherwise he will not account for the history of theories proper. And it will

not hurt the logician to remember that theories, far from being Platonic ideas, are created, reformed, discarded or forgotten by real people engaged in a historical process. Let us then approach the problem of theory construction.

2. Construction and Reality

2.1. *Theory Construction*

How are theories built? What are their building blocks and how are these put together? Are data good theory starters or does one begin with hypotheses? Are there any rules for building theories, so that a machine could be programmed to build them? How about theory reconstruction, in particular axiomatization: are there axiomatization kits? And what is the use of axiomatics anyway? These are some of the problem we shall tackle in the present section.

The building blocks of theories are propositions involving concepts of various degrees of remoteness from ordinary experience. Whereas modest theories are built with ordinary knowledge concepts, ambitious theories are built with either highly refined versions of the former or totally new concepts. Thus the concept of a set refines that of a finite collection, and the concept of a position coordinate that of place. On the other hand the concepts of tautology, vector space, electric field, protein synthesis, and elasticity of demand have no ordinary knowledge antecedents.

Whatever its origin and historical evolution, a concept belonging to a theory may be called *theoretical* if it is either peculiar to the theory or considerably elucidated by it. All of the concepts occurring in logical and mathematical theories are theoretical. And, because all the theories in advanced science and technology contain (usually voluminous) mathematical formalisms (or presuppositions), they contain theoretical concepts galore—in fact infinitely many in the case of any theory presupposing ordinary arithmetic. Some of the concepts of a theory are borrowed from its formal (logical or mathematical) background, whereas others are specific to it. Thus, in the theory of partial order we reviewed in Section 1.2 the concept of implication is taken from logic, whereas the order relation is peculiar to the theory.

The specific or technical concepts of a theory are in turn partitioned into *basic* (or undefined or primitive) and *derived* (or defined). Thus in the theory of partial order "$\leq$" is basic whereas "$<$" is derived. All definitions are

intratheoretical statements; more precisely, if explicit they are identities. (Vol. 2, Ch. 10, Section 2.2.) One adds them as needed. Occasionally one needs infinitely many. For example, in elementary number theory every number greater than 0 is defined as the successor of some other number, and so an infinitely long chain of definitions is generated.

The propositions (formulas) in a theory are either premises (postulates), definitions, or consequences of premises or definitions, i.e. theorems. (Recall Section 1.2.) The premises of a formal theory are tentative during the construction period, but cease to be so once the theory has been built and shown to be capable of solving some problems. From here on those premises are stipulations that determine the basic properties of the concepts of the theory: they create them. (Definitions proper, such as "2 equals the successor of 1", are noncreative. But axiomatic definitions, such as that of a partially ordered set, are creative.) The worth of the postulates of a formal theory lies exclusively in their power to create or refine concepts, unify previously scattered formulas, and entail new ones. Any conjectures that may arise in the course of investigating a formal theory are to be proved or disproved. When proved they become incorporated into the theory as theorems, when disproved they are dumped into the dustbin.

Having invented and organized a system of postulates and definitions is important but insufficient to solve any particular problems. To tackle the latter one must enrich that system with subsidiary assumptions. This work of specification or enrichment can be endless: witness the development of group theory, classical mechanics, and quantum mechanics. In all these cases the basic equations are adjoined now one set of subsidiary assumptions, now another, to yield in each case a particular theory capable of handling the given problem or allowing one to pose new problems. Thus in the case of group theory one may add the assumptions that the basic set is finite or else infinite, that the group operation is commutative or anticommutative, and so on. In the case of classical mechanics one may add that the thing of interest is a system of particles or a continuum, that the force is central or noncentral, that the viscosity is low or high, and so forth. And in the case of quantum mechanics one may add that the system is an atom or a molecule, that it is embedded in a magnetic field or an electric field, and so on. In this way intratheoretical progress is made, consisting in constructing special theories (bound models) on the basis of the given general theory. In other cases progress is accomplished by discovering the general theory underlying two or more special and formerly unrelated theories. In still other cases the contribution consists in transforming an

untidy theory into an axiomatic system. In all these cases advances are made with only pencil and paper.

The case of factual theories is radically different from that of formal theories. Here every premise that is not a definition is tentative, i.e. a hypothesis not a stipulation. It is supposed to prove its worth, jointly with the other premises, not only by its clarifying, unifying and deductive powers, but also by empirical tests suggesting that it is adequate to facts, i.e. factually true, to some extent. Sometimes the special hypotheses contained in theories are called 'data'—e.g. that a system is composed of so many things of such and such kinds, or that it is confined within such and such boundaries. Actually these are not data but corrigible assumptions which, together with the other premises, sketch the properties of the referent. Data have no place in the foundations of a theory, although every factual theory should make room for them. Thus in the theory of exponential growth in Section 1.2 the initial value $F(0)$ of the property in question, and its relative growth rate k, were left indeterminate, for they are to be determined empirically. Once we get such values we can insert them into the theory and use the latter to calculate, say, the value of F at any desired time t. (I.e. we add to the theory statements of the form "$F(0) = a$" and "$k = b$", where a and b are definite positive numbers.)

There are three reasons for excluding data from the foundations of a theory. One is generality. Thus mechanics is supposed to refer to all possible bodies of a certain kind in all possible (lawful) states of motion. This would not be the case if it contained detailed information about the masses and trajectories of the bodies. The second reason is testability: if a theory contained all the available information it would predict its own evidence and would therefore be untestable. The third reason is predictability: if we could know by sheer experience everything there is to know about things of a kind we would not need theories about them. In sum, the foundations of factual theories do not contain data. They only contain blanks to be filled by data.

Yet there is near consensus among scientists that every factual theory is or ought to be "based on" data such as measurement values. (They have not learned this from their research experience but in the classroom and in textbooks.) But it is not clear what 'based on' means in this context. The consensus view is so ambiguous that it can signify either of the five following theses: factual theories rest on data, they are built out of data, they summarize and extrapolate data, data motivate the construction of theories, or a theory, if true, is supported (confirmed) by data. The last two

claims are obviously true. Data, at least those of a new kind, call for explanation, and scientific explanation is best performed with the help of theories. And no theory can be pronounced to be a (sufficiently) true representation of a domain of facts unless it has passed some empirical tests. But the first three claims are false. Let us find out why.

First, as we saw a while ago, data do not belong to the foundation or basis of any theory. This is not a historical accident but a matter of logic. Indeed, every datum is a singular proposition, and singular propositions have no deductive power or, if preferred, they have too much of it, for they point to no hypotheses in particular. More precisely, a datum e implies an arbitrary proposition of the form "If h then e". (In fact, "$h \Rightarrow e$" is false only in case h is true and e false, which contradicts the assumption that e is true.) Second, most of the interesting properties of things, and all of the essential ones, can be observed or measured only with the help of theories, some of which are involved in the very design of measurement instruments. (For example, a way of measuring the mass of a body is to find its weight and divide the latter by the acceleration of gravity, which in turn is measured with the help of a pendulum and the theory of the pendulum. See Figure 9.7.) Third, no measurement can yield any of the sharp values of any of the

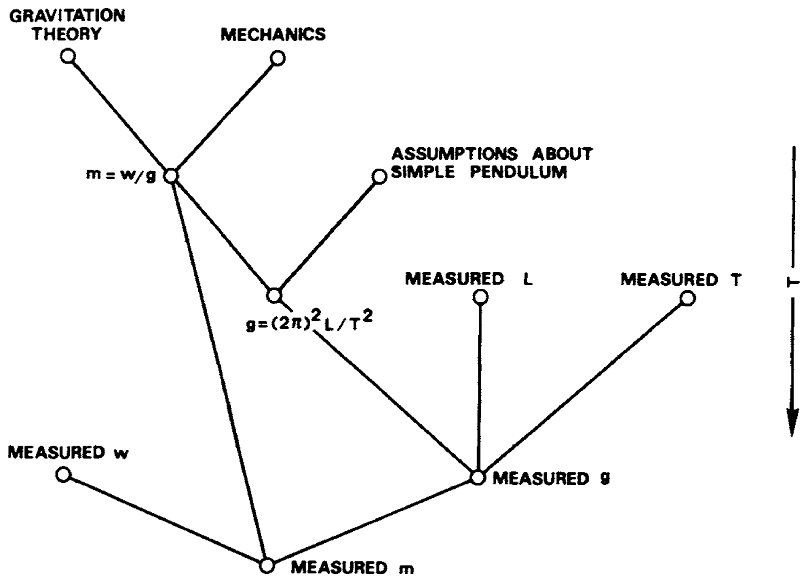


Fig. 9.7. The measurement of mass with the help of a scale (for measuring weight w) and a pendulum (for measuring acceleration of gravity g).

continuous (or piece-wise continuous) functions that occur in all of the advanced factual theories. In particular, the instantaneous values of rates (e.g. speeds) can be calculated but not measured exactly, for measurement yields only averages. (Thus we calculate the rate of change of a property represented by a function F by computing its time derivative, but measurement gives only values of the ratio $\Delta F/\Delta t$ of finite increments.

The problem of whether theories can be built out of data can be approached in the light of the distinction between direct and inverse problems (Ch. 7, Section 4.1). To explain or forecast data with the help of a theory and prior data is a direct problem with a unique (though possibly many-membered) solution. The corresponding inverse problem is: Given a set of data build a theory covering them. Like most inverse problems, this one is indeterminate, both in the sense that data are insufficient, and that there is always more than one theory that explains any given set of data. (Philosophers call this well-known fact "Quine's thesis of indeterminacy of translation".)

Consider the following inverse problem: Given (by observation or by conjecture) the genetic structure of a population at a given time, and assuming a constant environment, find its genetic structure (the distributions of the genes) after a certain number of generations. This is a direct problem that can be solved with the theory of population genetics, which contains precise formulas for simple cases. The inverse problem, of finding such formulas from a study of populations in the wild or in the laboratory, is hopeless, not only because nature does not bother with the simplifying assumptions the theorist is forced to make, but also for the following reason. Most populations in nature have attained a state of equilibrium, i.e. a constant genetic composition, and such equilibrium states can be reached from many different initial states, much as a pebble can reach a given place at the foot of a mountain from many places on the latter. So, the inverse problem is insoluble. Data do not entail theory but may motivate its construction, and theory does not entail data but may help gather them. Only the merger of theory and data can entail further data. Recall Figure 9.7.

If data do not trigger theorizing, what can? Hypotheses can. The theorist usually starts with a handful of more or less disconnected generalizations and, of course, a bag full of mathematical tricks. Some of those generalizations are nothing but empirical generalizations, such as curves that interpolate or extrapolate observational data; others are speculative conjectures. And among the latter some may be mere programmatic

conjectures, or schemata to be spelled out, such as "A is some function of B". These trigger hypotheses may be called the *precursors* of the theory.

Building a system out of precursor hypotheses is not merely a matter of collecting them. They have to be refined and interconnected, possibly by further hypotheses. The refinement may be performed with the help of mathematical tools, and the glue hypotheses may contain concepts not occurring in the precursors. A classical yet often misunderstood example is Newton's creation of classical mechanics. His precursor hypotheses were the kinematical laws conjectured and checked by Galileo, Kepler, Huyghens, and a few others. Newton could not generalize these laws into a general kinematical law because there *is* no such law. What he did was to invent new, dynamical hypotheses containing the new basic concepts of mass and force, to which that of stress was added later on. It turned out that Newton's theory contained not only the precursor hypotheses but also any number of new ones, one for each possible force law, mass distribution, and stress configuration. (Newton's laws of motion are actually schemata that can be specialized by specifying the forces, masses, and stresses.) Likewise the mathematical models that are emerging in neuropsychology contain neurophysiological concepts, such as those of neuronal connectivity and plasticity, and aim at relating and explaining the regularities found by the students of behavior and cognition, as well as many unsuspected ones.

It is difficult to exaggerate the importance of mathematics in the construction of scientific and technological theories. We take mathematics so much for granted nowadays that we sometimes forget the ways in which it helps theory construction. For one thing mathematics offers a huge store of ready made concepts and formulas that, to speak anthropomorphically, are only waiting to be used and interpreted in factual terms. So much so that every factual theory in advanced science or technology is a mathematical formalism together with a factual interpretation of some of its concepts. (See Section 2.2.) For another, mathematics offers a large supply of methods for solving, exactly or approximately, nearly all the problems that are formulated mathematically. So much so that formulating a problem in mathematical terms is usually the hardest stage of a research project in theoretical science, though solving it may be the most time-consuming. (Computers are a great help in solving problems of certain kinds, not in formulating them.) Thirdly, mathematics guarantees exactness and deductive power, and facilitates the gluing and supplementing of the precursor hypotheses of a theory. So much so that the theorist who employs mathematics feels that he is being led by the hand and that his

thoughts spread over an entire network that seems not to be his, to the point that he meets unexpected problems and surprising solutions. For a simple mathematical model see Appendix 1.

In the process of building a scientific theory or model with the help of mathematics we can distinguish the following stages:

(i) *survey* of the relevant available knowledge with a vew to finding candidates to precursor hypotheses;
(ii) *invention* of stronger hypotheses, usually containing sophisticated transempirical (nonobservational) concepts, that may entail the precursors;
(iii) *checking* whether the new hypotheses entail the precursor hypotheses or close enough versions of them;
(iv) *mathematization* of the strong hypotheses—often but an aspect of invention, and occasionally requiring the forging of new mathematical ideas;
(v) *investigation of the mathematical problem* with momentary oblivion of the possible factual interpretation—to the point that this stage of research is sometimes taken up by mathematicians;
(vi) *simplifications* facilitating the handling of the mathematical problem or even rendering it tractable;
(vii) *factual interpretation* of some of the results of the previous investigation.

All of these seven steps are equally important. In particular, it is not true that one can always get by without mathematization, or that the essential feature of modeling is simplification. Nonmathematical thinking is clumsy and laborious, inexact and rarely systematic, to the point that purely verbal (i.e. nonmathematical) theories are at best good precursors of mathematical theories. With the help of mathematics we "see" systems, patterns and consequences where none are apparent without it. As for simplification, it is true that it is inherent in theorizing (and also in classification), but it is not more important than invention: the beard must be there before it can be trimmed. And in any case simplicity, if found inconsistent with truth, can be remedied by inventing further concepts or propositions. After all, the historical process has not been that of theoretical simplification but, on the whole, one of progressive complication. Only, such complication is not arbitrary and is not introduced to save some dogma—as is the case with

ideological complication—but is motivated by the wish to perfect agreement with fact or a deeper understanding of fact. (See Bunge, 1963.)

In dealing with theories we must avoid several pitfalls. One of them is the very natural attempt to handle new problems with the help of existing theories. While this strategy works in some cases, it is bound to fail in others. Understandably, such conservatism is more likely to succeed in the advanced branches of science and technology. Thus an engineer would be ill advised to try and build a new electrodynamic theory every time he faces a new problem in power transmission or in electromagnetic communication. On the other hand psychologists and philosophers are wrong in insisting that the central nervous system is nothing but a Turing machine, for automata theory does not even contain time, which is essential to account for the variability of neural connections.

A second pitfall to be avoided is that of mathematizing the wrong precursor hypotheses, and embellishing theories that are known to be false or at least are not known to be even approximately true. Contemporary microeconomics, with its powerful mathematics and its slender empirical basis, is a case in point. What is the point of embellishing the neoclassical hypotheses concerning consumer behavior and the way prices are fixed if they are at best untested, at worse at variance with empirical research? Better construct modest econometric models than waste effort studying nonexisting equilibria. (More in Vol. 6.)

A third pitfall is procrastination in starting to theorize for fear that it may be premature. This fear is unjustified, for (*a*) theories, even if false, advance experimental research by suggesting new problems or new ways of approaching known problems; (*b*) knowledge is not properly scientific unless it has a strong theoretical component; (*c*) accumulating empirical information in the absence of theories may prove wasteful, for research that fails to be inspired by theories is bound to yield at best superficial and disconnected data, and at worse irrelevant ones. It is never too soon to start theorizing—and never too late for changing a defective theory for a better one.

Having decided that the time has come to build a theory, can we ask methodology for directions? Is there a logic or technique of theory construction? The answer is in the negative: there are only philosophical fantasies on them. The very idea that original work could be rule-directed is self-contradictory, for originality defies the known rules. (Instead of speculating about the existence of a logic of theory building, we should speculate about the possible neural mechanisms of theory invention, as well

as about the social mechanisms that favor or discourage the invention of theories.) Since invention cannot be mechanized, no machine could be programmed to build theories for us. Machines can help once we have created theories and algorithms—e.g. they can help solve some mathematical problems that arise in applying the theories. But even here machines can do only so much. For example, they cannot handle continuity. And their output, namely a bunch of printouts, is far more difficult to interpret than even the least accurate of analytic solutions.

The case of theory reconstruction, or axiomatization, is different: here there is a list of definite do's. (Cf. Bunge, 1973a.) Thus the first step is to investigate the theory to be axiomatized with the aim of spotting the most likely candidates for undefined (primitive) concepts and unproved formulas (postulates)—i.e. the strongest constructs. A second task is to define (implicitly) every undefined notion by means of a set of postulates characterizing both its mathematical form and its factual content if any. (E.g. "F is a piece-wise continuous vector-valued function on the cartesian product of A by B", and "The value of F at $\langle a, b\rangle$, where a is in A and b in B, represents the F-ity of thing a at place b".) Thirdly, we should check whether the proposed system of postulates entails all of the standard theorems in the theory. If not, we should promote some other formula(s) to the axiom rank, or invent an even stronger formula. There are a few other heuristic rules we can use to facilitate axiomatization, but none of them guides the creation of a new theory *ab initio*. In this task only other people, in particular mathematicians, can help the creative theorist.

(It may be objected that modern metamathematics has shown the limits of axiomatizability, particularly with regard to completeness and decidability. For example, the following result has been proved for an arbitrary theory T formalized within the first order predicate calculus with identity: *If T is complete, then T is not axiomatizable iff T is undecidable*: see Tarski *et al.* (1953). Since the vast majority of theories are undecidable, it would seem that axiomatization is the exception rather than the rule. This is not true, for the theorem holds for *complete* theories, and in science and technology there are no complete theories. Moreover, in these fields we have no use for complete theories, as a theory can be applied only if it is *incomplete*, for only then can it be enriched, without contradiction, with subsidiary hypotheses and data. So, the above metamathematical result has no bearing on the axiomatizability of factual theories.)

Should it be asked what the value of axiomatics is, we could answer with another question, namely 'What is the value of cleanliness and tidiness?' In

axiomatizing a theory we are bound to discover gaps or inconsistencies, refine concepts, overhaul the mathematical formalism, and reorganize the formulas in a clearer and more reasonable fashion. In the process we may discover new theorems or discover that some formulas that used to be taken for theorems are not really part of the theory. We may facilitate the spotting of the deficient presuppositions or hypotheses to be corrected. We may establish that certain formulas are definitions, and thus save the experimentalist the trouble of trying to test them. We may dispel any doubts concerning the genuine referents of the theory. Also, an axiomatized theory is easier to recall and therefore use. Finally, axiomatization renders a deep and cogent philosophical analysis of theories possible.

Unfortunately a romantic wind is blowing in philosophy once again. Romantics abhor formalization and even analysis, preferring to make sweeping, global, vague, and sometimes derogatory assertions about theories, theorists, and society. There seem to be several causes for this revolt. One is that it is but one aspect of the general movement against science and technology that accompanies both the revival of religious fundamentalism and the youthful rebellion against the establishment. Another is that most of the philosophical formalizations in the past, notably those proposed by Carnap and Reichenbach, turned out to be unfaithful to science and philosophically unilluminating. A third cause is that it is far easier to revolt against formalism than to learn it. This revolt is unreasonable. If one type of formalization does not work, e.g. because it bears on cartoons of scientific theories, or on trivial problems, then one should try a better one. And if one dislikes formalism, one should stay clear of advanced science and technology, which carry mathematics in their very bones.

To conclude, there is no method for building theories other than a command of the basic facts, the founder generalizations, and some mathematical tools. The so-called inductive method does not deliver theories because data collection itself is done in the light of some conjecture or other. Nor does the so-called deductive method yield theories, for it works only where the initial hypotheses have already been conceived. Theory construction is not a rule-directed process but an original brain process to be studied by psychology and social science, not by logic. Such an empirical study of theory construction must pay particular attention to the background knowledge and values of the theorist, to the way he sees the world and our knowledge of it. (See Holton (1973) for the importance of general guiding principles in theorizing.)

2.2. *Theory and Reality*

We have called a theory *formal* if all its referents are constructs, and *factual* otherwise (Section 1.2). Formal theories are then, by definition, unrelated to reality—except of course that they are invented, learned, and used by real beings. On the other hand factual theories are doubly related to reality: semantically and methodologically. In fact they *refer* to certified or putatively real things, and they are *tested* by confrontation with experience, which is a part of reality. In this section we shall be concerned with the semantical relation of representation, leaving the methodological relation of confirmation for Ch. 12, Section 2.1.

Because a factual theory is about certified or putatively real objects, it is more than a formalism: it also includes propositions telling us what refers to what and what represents what. We call these propositions *semantic assumptions*; in the literature they are usually misnamed 'correspondence rules' or even 'operational definitions'. In our view semantic assumptions are corrigible hypotheses, not rules or definitions: we assume them and then check whether the theory containing them passes some empirical tests. If it does, then the system composed of the formalism and the semantic assumptions is pronounced true until further notice. Otherwise we introduce changes in the formalism or the semantic assumption—or abandon the theory altogether. For example, a mathematical model originally designed to account for the propagation of a contagious disease may not do this job but may instead represent adequately the propagation of rumors: the formalism is kept and only the semantic assumptions are modified. See Appendix 2 for an example.

Every semantic assumption is an *interpretation*, i.e. a formula of the form "Concept X stands for (or represents) Y". Some interpretations assign constructs to constructs, others to factual items. We call the former *formal*, the latter *factual* interpretations. Thus the basic (undefined) concepts of the theory of partial order (Section 1.2) can be interpreted as follows:

$$\rho(R) = \textit{The set of natural numbers},$$
$$\rho(\leq) = \textit{Less than or equal to}.$$

This interpretation yields a *model* or example of the abstract theory of partial order: i.e. the formulas of the latter are satisfied under that interpretation. This model does not represent any facts: its semantic assumptions relate concepts to concepts. If, on the other hand, we set

$$\rho(R) = \textit{Collection of lamp posts along a highway},$$
$$\rho(\leq) = \textit{To the left of},$$

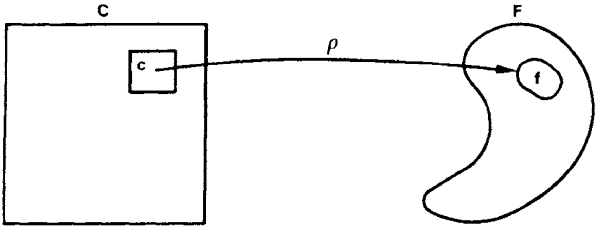


Fig. 9.8 The representation function ρ assigns a subset c of the set C of constructs a subset f of the collection F of factual items.

we produce a model of a collection of factual items. We say that the factual interpretation has produced a *factual model*, i.e. a theory representing a collection of factual items: See Figure 9.8.

The expression 'factual model' is ambiguous, for it signifies an example of an abstract theory and also a model of some factual domain. In the above case the factual interpretation of R and $\leq$ produced a model that is both. Likewise a linguistic interpretation of the theory of semigroups, and an interpretation of Boolean algebra in terms of switching circuits, yield models in both senses. However, not every factual model is of this kind. Most models in science and technology involve mathematical formalisms that have already been interpreted in mathematical terms—e.g. not abstract geometry but Euclidean geometry, not the general theory of functional spaces but the theory of Hilbert spaces. See Figure 9.9. However, the essential point is that factual models, unlike formal models, represent certified or putatively real factual items—material things, their properties, and changes in the latter. Because of this factual reference they must be tested not only conceptually but also empirically.

(Whether conceptual or factual, an interpretation may be construed as a function $\rho: 2^C \rightarrow 2^D$ assigning to every subset of a set C of constructs a subset of a different collection D. If D is taken to be a set of nonabstract but still conceptual items, the interpretation is called *formal*, as when we assign to every subset of C a set of natural numbers, or of points on a sphere, or of real valued functions. Model theory, or the semantics of mathematics, handles only formal interpretations. On the other hand the semantics in Vols. 1 and 2 of this *Treatise* focuses on the case when D is a set of factual items, such as things, properties of things, or events in things. In this case ρ is called a *factual* interpretation, and the corresponding model a *factual* one. Actually factual interpretations are *partial* rather than full functions, for they do not assign to every construct a collection of factual items. Thus a theory presupposing the infinitesimal calculus may contain all the infinitely

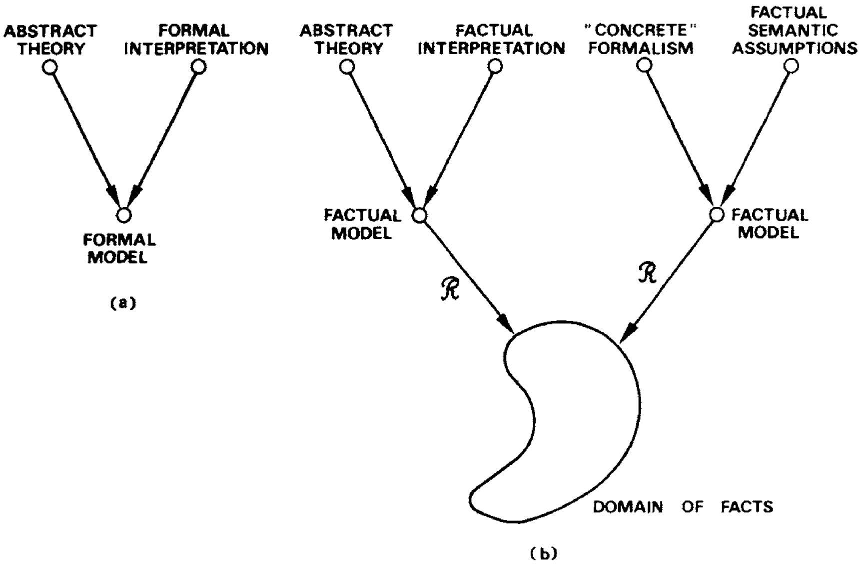


Fig. 9.9. Two types of model. (*a*) Conceptual models within formal science: no facts are being represented. (*b*) Conceptual models within factual science and technology: whereas some result from interpreting abstract theories, the great majority result from superimposing factual semantic assumptions on mathematical formalisms that have already been interpreted in mathematical terms—e.g. where sets are not abstract or generic but are composed of specific objects such as real numbers or points in a Euclidean three-space.

many derivatives of every function occurring in the theory, but only the first few derivatives are assigned factual interpretations, e.g. as velocities or accelerations.)

(In view of the preceding we may now characterize theories in a fuller way. An *abstract theory* is, as stated in Section 1.2, merely a logically organized context, and so characterizable as a quadruple $\mathcal{T}_a = \langle P, Q, R, \vdash \rangle$. On the other hand an *interpreted theory* is a logically organized context some components of which are assigned an interpretation, i.e. $\mathcal{T}_i = \langle P, Q, R, \vdash, \rho \rangle$, where $\rho: 2^C \to 2^R$, and $C = P \cup Q$. In particular, a factually interpreted theory, or *factual theory* for short, is an interpreted theory some components of which are assigned a factual interpretation, i.e. $\mathcal{T}_f = \langle P, Q, R, F, \vdash, \rho \rangle$, where now R includes a nonempty collection F of factual items, and $D = F$, so that $\rho: 2^C \to 2^F$ is a partial function assigning factual items to some constructs in $C = P \cup Q$.)

The notion of factual interpretation allows us to analyze the set P of propositions of a factual theory $\mathcal{T}_f$. To simplify the discussion, assume that

the theory has been axiomatized. Then the set of its postulates is divided into two disjoint subsets: the set M of all *mathematical* axioms, and the set S of all *semantic* assumptions. Every mathematical assumption is a formula of some mathematical theory, and every semantic assumption is of either of the following types:

For sets: "Set A equals such and such kind of (presumably) real entities."

For functions: "Function F represents such and such property of entities of kind K."

A first consequence is that the totality P of propositions of $\mathcal{T}_f$ equals the set of consequences of the union of M and S, i.e. $P = Cn(M \cup S)$. A second consequence is that the set P of formulas of $\mathcal{T}_f$ is divided into two disjoint subsets: (*a*) the set P_m of purely mathematical formulas, i.e. those specifying formal properties of the predicates, and (*b*) the set P_f of formulas that are assigned a factual interpretation by virtue of the factual semantic assumptions of the theory. For example,the propositions "A is an algebra of sets" and "F is twice differentiable" are purely mathematical formulas. On the other hand the law statements and the definitions containing factually meaningful concepts, such as the definition of zero centigrade, belong to the set P_f of factually meaningful formulas.

Let us insist that the semantic assumptions are necessary components of all factual theories because a mathematical concept can be interpreted in alternative ways or in none. For example, an increasing function may represent an expanding volume in one theory and a growing population in another. Also, the semantic assumptions are not conventions like the definitions: they are hypotheses and as such they may prove false. For example, in physics an interpretation of an equation as describing a process is said to be 'unphysical' if it allows an event to occur before its cause; and in biology a semantic assumption interpreting a certain function as the total population of an ecosystem is dubbed 'unbiological' if that function takes on negative values.

That the semantic assumptions are hypotheses does not mean that they can be checked empirically in isolation from the formulas they are attached to. What are testable are these formulas together with their factual interpretation or, rather, the factually interpreted formulas. For example, the formula "$y = ax + 1$", where every symbol designates a positive real number, is not empirically testable. On the other hand the following is: "$y = ax + 1$, where 'y' represents the ratio of lengths of a metal bar at temperatures x and 0, and a is the expansion coefficient of that particular metal".

Any given thing can be represented by different theories. This is not a datum but an epistemological hypothesis confirmed but not proved by historical research. It is an optimistic hypothesis that encourages theorizing and moreover it is an irrefutable hypothesis. Indeed, were we to face things of some kind that appeared to resist theoretical modeling we would rather blame our own incompetence than the things themselves, and we would hope that others should succeed where we have failed.

Factual theories represent their referents but are not perfect pictures of them. They are not pictures at all but symbolic constructions bearing no resemblance to the objects they represent. (Boltzmann wrote that the relation of theories to reality is like that of letters to spoken words, and Einstein that it resembles the relation between numerals and coats in the cloak room.) Nor can theoretical representations be perfectly faithful: they are at best approximately true. Hence naive realism is just as wrong as conventionalism. Critical realism is the ticket.

The different theories representing things of a given kind are not necessarily equivalent: some may be deeper or truer than others. For example, consider the following sequence of models of the terrestrial atmosphere that have been considered historically, and are still being studied:

1. Flat earth, homogeneous film of dry air, no radiation.
2. Round earth, homogeneous film of dry air, no radiation.
3. Flat earth, different strata of dry air, no radiation.
4. Round earth, same.
5. Flat earth, strata of moist air, no radiation.
6. Round earth, same.
7. Flat earth, each stratum of moist air divided into homogeneous parallelepipeds, no radiation.
8. Round earth, same.
9. Flat earth, same plus radiation.
10. Round earth, same.

Each model is more realistic than the preceding in at least one respect, and the last is the most realistic (the truest) of all. But even the first model, though very coarse, has some use, for it allows for winds—although, without radiation, it cannot explain them, and without moisture, it makes no room for precipitation.

This concludes our accout of the semantics of factual theories. (For a more detailed treatment see Vols. 1 and 2.) Our account differs radically

from the view commonly held by mathematicians or philosophers, which neglects the factual semantic assumptions. One such popular view is *Pythagoreanism*, according to which things *are* constructs instead of being represented by constructs. Thus some writers state that a particle is an oriented one-manifold, an electromagnetic field a tensor field, and in general a physical system "a set of individuals together with numerical functions on these individuals" (Sneed, 1979). This view is incapable of distinguishing between mathematics and factual science. It does not explain how an unchanging construct, such as a tensor field, can be identical to a changing thing, such as an electromagnetic wave. Besides, it is contradictory, for it identifies a given construct now with one thing, now with another, instead of stating that the construct represents one thing in one theory, and another thing in another.

A related view is *formalism*, according to which model theory, i.e. the semantics of mathematics, suffices to analyze factual theories (Sneed, 1979; Stegmüller, 1976). The upholders of this view have probably been misled by the ambiguity of the term 'model', which in metamathematics designates an example of an abstract theory, whereas in science and technology it designates a specific theory (Section 1.2). Consequently they ignore the concepts of reference and representation, central to the semantics of factual theories, and they cannot explain why factual theories cannot be validated or invalidated by purely mathematical considerations. (For further criticisms of Pythagoreanism and formalism see Vol. 2, Ch. 6, Section 3.)

Finally there is *operationism*, or the doctrine that factual theories are mathematical formalisms enriched with operational "definitions", i.e. descriptions of the way the properties represented by the theory are measured. This view is far closer to the truth than the two previous views, but it confuses matters of reference and representation with questions of test. It ignores that one and the same property may, in principle, be measured by different methods. (For further criticisms of operationism see Vol. 1, Ch. 5, Section 3.4, and Bunge, 1973a, Ch. 1.)

In contrast to the above views ours is realist and it has been incorporated into the axiomatization of several basic physical theories (Bunge, 1967b).

3. Convention and Law

3.1. *Convention*

There are several kinds of convention, and every theory contains at least conventions on nomenclature. Some conventions, like counting the

negative numbers on the left side of an axis, are mere customs (that might not have been adopted if we all were left-handed). Others, like definitions, are time-saving and clarity-gaining devices. Some conventions are dispensable and all are replaceable. Also, some conventions are more useful than others. For example, in daily life the meter is handier than either the fermi, convenient in atomic physics, or the light year, convenient in astronomy.

We shall study conventions of the following kinds: notational, definitions, conventions about scales, and simplifying assumptions. *Notational conventions* assign nonlinguistic objects to signs. They are of two kinds: designation rules and denotation rules (Vol. 1, Ch. 1). A *designation rule*, such as "'$\mathbb{N}$' designates the set of natural numbers", assigns a construct to a symbol. (Nominalists do not need designation rules and they have no concept and no name for that which remains invariant under changes of name.) And a *denotation rule*, such as "'W' denotes the direction in which sunset occurs", assigns a factual item to a symbol. Similar conventions are those concerning the positive direction of rotations, the sign of electric charges, and the choice of the initial and final stage of a process. Every single notational convention is replaceable, though perhaps at the risk of confusion. But every explicit context, in particular every theory, should contain a set of notational conventions—unless they are notorious.

A *definition* is a horse of a different color: it stipulates the identity of two concepts or of their symbols. Thus we may define "positron" as "positive electron" and "1" as the successor of 0 in the set of natural numbers. Such identities are adopted by convention, whereas others, such as "For all real numbers $x, (-x)^2 = x^2$", are provable. The specific difference between a definition and an identity is then that the former is stipulated. Such difference is strictly methodological. On the other hand the identities "Heat = Kinetic energy of randomly moving atoms or molecules", "Gene = Segment of DNA molecule", and "Mental state = State of a plastic neural system in a living brain", are hypotheses to be justified empirically. They state identities of factual items, not of concepts; they state the synonymy of different expressions.

Whether or not a formula is a definition depends on the context in which it occurs. Thus whereas in elementary mechanics forces are undefined and linear momenta defined, in Hamiltonian dynamics forces are defined and linear momenta are primitive. And depending on the context we may postulate or prove the existence of other minds. (In principle it can be proved from the basic similarity of all human brains.) The contextuality of

definitions is just a particular case of the contextuality of concepts. Thus the concept of mammal does not occur in pure mathematics, and that of creation out of nothing is out of every scientific context. Only mathematical and philosophical concepts can cross all disciplinary borders.

That definitions are conventions does not mean that they are wholly arbitrary. In a well-organized context the place of a definition cannot be shifted without reorganizing the entire system, and the defined concepts cannot exchange places with the defining ones. For example, numbers can be defined in terms of sets but not conversely. Likewise, "process" can be defined in terms of "thing" but—*pace* Whitehead and his followers—the converse is impossible. (Indeed the very notion of a process, or sequence of states in a thing, involves some concept of a thing. Therefore the definition of "thing" as "a bundle of processes" is circular.) Given a choice of undefined concepts, the strongest ones are preferable, and the defined ones cannot be introduced haphazardly but in an orderly fashion. (Thus the definition of state of a thing presupposes the concepts of thing and of property of a thing.) Moreover, definitions are sometimes backed by axioms or theorems, so they cannot be introduced until such formulas have been stated. Thus the notion of an inertial reference frame is definable in terms of equations of motion, or of field equations, which in turn presuppose the general concept of a reference frame, as well as the concepts of space, time, and much else.

Definitions relate concepts but this does not mean that they lack factual reference or factual presuppositions. For example, the definition of "100° C" as the (same as) the boiling point of pure water at sea level has a factual content: it refers to water. And the definition of "machine" as "artifact for transforming energy of some kind to some purpose" presupposes the law of conservation of energy. Indeed a device alleged to create or annihilate energy would not be called a machine.

Even if a definition has an external reference, it does not introduce new knowledge and cannot take the place of a hypothesis. (This does not hold for axiomatic definitions, which are creative and, in the case of science, are systems of hypotheses.) In particular, a definition belonging to a theory in pure mathematics does not describe any processes in the real world, not even if interpreted in factual terms. (This is ignored by the Bayesian school of epistemology, according to which Bayes' theorem represents the way we learn from experience. This is false, for the theorem is an immediate consequence of the definition of conditional probability.)

Since definitions define only concepts, they cannot define things. This

must be stressed because some philosophers believe that the referents of factual theories can be defined the same way concepts are defined, e.g. in terms of a system of equations. What such a system does is to *describe* its referents not to define them. What is true is that a theory may help *identify* factual items as members of a class. For example, Newton's mechanics, together with certain measurements, helps us identify bodies, genetics genes, and so on. In short, an adequate (sufficiently true) theory (*a*) describes its referents, (*b*) defines some of its own concepts, and (*c*) is a criterion for identifying (not defining) its referents. On the other hand most definitions are not utilizable as criteria. Thus a definition of consistency is no criterion for proving or disproving the consistency of a theory, and a definition of truth does not suffice to assign truth values.

All factual theories, and all measurements, involve *units*: of length, time, mass, etc. Units are conventional and they are involved in the very characterization of magnitudes such as physical quantities. Thus when giving the value of the mass of a body b relative to a reference frame f, it is essential to mention the unit u in which it has been measured or computed. In short, such a datum or assumption has the form "$M(b, f, u) = m$", where m is a positive real number. (In general, the property mass is represented by a function of the form $M: B \times F \times U_M \to \mathbb{R}^+$, where B is the set of all bodies, F that of all reference frames, and U_M that of all conceivable mass units, whereas $\mathbb{R}^+$ is the set of positive real numbers. See Bunge (1971).)

The meter, the second and the kilogram can be taken as fundamental units; the micron, the year and the gram as derived units. Units are conventional but not wholly arbitrary. Thus the second, or more precisely the mean solar second, is defined as 1/86,400 mean solar day, which in turn results from astronomical observations and calculations. The fundamental units are materialized as standards, which are reproducible in any well equipped laboratory. And such standards, far from being arbitrary, are chosen because of their accuracy, stability and reproducibility. A good example is the atomic length standard, defined as the wavelength of a certain sharp spectral line that is easy to prepare and measure, and that has a known theoretical description (namely as the result of a certain atomic transition).

A full homogeneous scale is equipped not only with a unit but also with a starting point or *zero*. The absolute or Kelvin zero of temperature is one example. The year 1 A.D. is another such conventional point of departure—the number 0 having been unknown to the makers of the Western calendar. Such choices may look natural and in a way they are, but

they are conventional. Thus we could shift the zero of temperature below its current position by an arbitrary amount. However, the additional stretch would have no use, since atomic and molecular motion cease at the universally accepted absolute zero. (As a consequence we would have to subtract a certain constant from all the energy values.)

Next in our list of conventions come *simplifying assumptions*. All except fundamental theories contain some simplifying assumptions. For example, in a certain calculation or measurement we may pretend that the value of *pi* is 3.14; that the system of interest has no internal structure, or is perfectly well isolated, or has only the properties we happen to know; that certain discontinuous variables, such as population, are continuous—or conversely; that brain functions are not affected by social circumstances; that unemployment or inflation has a single cause, etc. All such simplifying assumptions are introduced to expedite modeling or inference, or even to make them possible. They are white lies, for they serve the search for truth. They become black lies when presented as truths or when we keep them even after having been shown to be wide of the mark—as is nowadays the case with the assumption of perfect competition, still at the heart of most economic models.

In sum, all theories contain conventions and, paradoxically, factual theories contain the most. The epistemological moral is obvious: Although factual theories may represent their referents in a fairly true manner, such representations are not pictures or copies.

3.2. *Law*

Man is a compulsive pattern seeker. We abhor the idea of chaos and tend to see regularity even where there is none. Thus, unless we are familiar with probability, we may feel tempted to commit the gambler's fallacy, i.e. to expect that a streak of bad luck will be compensated for by one of good luck, and conversely. We expect order and this is why we look for it. But order is seldom perceptible: the phenomenal world, unlike the real world, is chaotic. Therefore pattern must be either hypothesized or imposed. We hypothesize pattern when we hazard a hypothesis about some constancy or some trend; we impose order when we enforce a rule of action, i.e. compel things to fall into step.

Let us examine the ideas of pattern, trend, empirical regularity, law, and rule. Any stable or reproducible configuration of objects of some kind may be called a *pattern* even if it is the outcome of chance. *Examples*: a string of

pebbles left on the sand by the tide, a finite sequence of tones, a behavior pattern. Common to all these instances are order and stability (or repeatability or reproducibility). Every pattern is some order in some collection of objects, and can thus be represented as a structured set. The structure is a set of relations, possibly functions. The converse if false: even a totally chaotic (but finite) set of points on a plane can be represented by a function. Only, such function will not belong to a family, will not be continuous, will not solve an equation, etc.: it will be an *ad hoc* or stray function. More on this below.

Patterns are stable for a while but need not be eternal. Internal tensions among the members of a system, or environmental changes, may end up in pattern changes. Thus a teacher, trainer or experimenter can teach an animal to adopt new behavior patterns by rewarding actions of certain kinds. But he can just as easily extinguish the newly acquired behavior pattern by negative reinforcement. In either case the teacher forces a pattern change, or repatterning. (He does so with the help of certain learning laws.) In the case of biosystems and sociosystems repatterning can occur without change in species. Thus a biopopulation occupying a new territory, and doing well in it, does not change species immediately. Likewise a social system surviving a drastic change, such as revolution or war, may still be what it used to be, e.g. a production center or a school. On the other hand, in the case of physical and chemical systems repatterning is always accompanied by changes in kind or species.

One of the most interesting problems is that of the emergence of pattern from the interplay of initially independent, hence randomly behaving, entities. For example, dust is distributed irregularly throughout the universe, but gravitational attraction clumps it; if several penduli are suspended from an elastic bar, they end up by oscillating synchronically regardless of their initial positions; any piece of cast metal is a random agglomeration of grains of irregular bulks and shapes, but when drawn into a wire or rolled out into a sheet they tend to adopt a common orientation; genic mutations and recombinations are random, but the environment makes certain uniform demands and thus produces some order; likewise the actions of persons are constrained by nature and society, and thus fall into patterns. In all such cases the scientific problem is to explain the emergence of pattern, or the repatterning, in terms of laws and circumstances. And the technological problem is to favor, channel, or prevent processes of pattern emergence and repatterning.

A pattern is called a *trend* when it is a uniform sequence of states of, or

events in, a system. Thus the expansion of the universe, and the increase in the prices of most commodities over the past decade, are trends. So are the increasing effectiveness of the hands along the evolution of primates, the steady increase of the total human population since the industrial revolution, the increasing literacy and the rising life expectancy in the Third World, and the increasing lethality of weapons.

Trends are stable while they last. They can stop or even reverse, particularly if the events concerned are brought under human control. Thus at the time of writing it is conceivable that, at some point in time, the universe will start contracting and prices will start falling. The uncovering of trends is an important task in every discipline. It is often the work of educated intuition, and sometimes the result of applying certain statistical techniques. (The whole of econometrics is an effort to discover and represent economic trends.) But we should not rest content with knowing about trends: we should try and explain them in terms of laws and circumstances.

Trends are not laws but they are often confused with laws in the life and social sciences. A good example is the so-called Lartet law, according to which the brains of mammals and birds have increased steadily in relative size along evolution. This is a typical trend, so much so that most of the species to which it applies have become extinct. Or take Marx's objective of discovering what he called "the laws of motion of the capitalist society". He did not attain it, but he did discover a number of trends or short term tendencies. Some of them, such as the increasing concentration of capital, are still in force; others, such as the increasing pauperization of the working class, have been reversed in the industrialized countries. (More on the difference between trends and laws in Popper (1957) and Bunge (1967a).)

Next in line come the empirical generalizations occurring in all branches of factual knowledge. A formula may be called an *empirical generalization* if it is a summary of data or an inductive generalization from data. In science and technology the typical empirical generalization is the smooth curve fitting a finite set of observational or experimental data: Figure 9.10. The empirical generalizations of this type are full fledged hypotheses, not data, for they interpolate between data: they concern not only actuals but also possibles. Consequently they are less, not more certain than the data they cover—so much so that every set of data, being finite, can be covered by infinitely many smooth curves passing near them. True, in the absence of a theory we usually choose the simplest such curves. However, simplicity is convention, not law. Indeed further research may well, and often does,

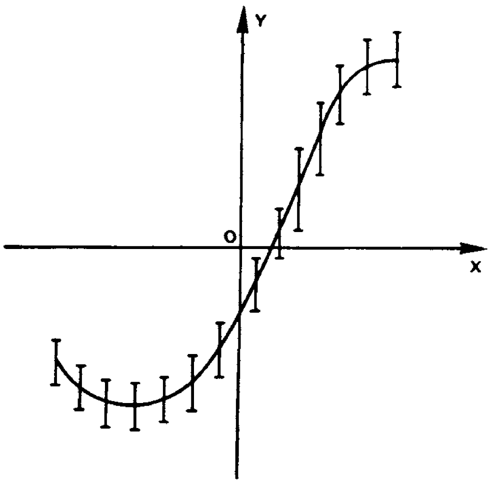


Fig. 9.10. One of the infinitely many possible curves passing through a set of data. The vertical segments, possibly unequal, represent the errors of observation.

refute the simplest hypothesis, forcing us to adopt a more complicated one exhibiting the fine structure of the factual domain in question. See Figure 9.11.

An empirical generalization is supported only by the data that suggested it: it is a stray not a systemic hypothesis. Should new unfavorable data appear, we will be forced to give it up, for we have nothing else to fall back on. If the generalization has been obtained by using a standard in-

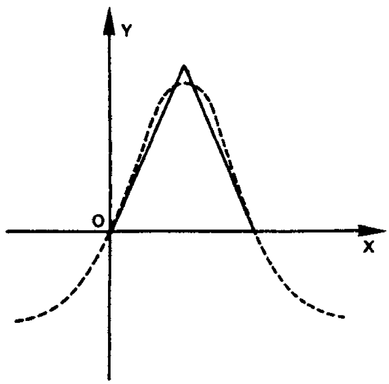


Fig. 9.11 Data suggest full line. Theoretical investigation suggests dotted line.

terpolation or curve fitting technique, it can easily be corrected by changing the values of some of the parameters occurring in it. But this methodological advantage has a high semantic price, for such parameters have seldom a factual meaning. So, paradoxically, empirical generalizations involve concepts without real counterparts. As we shall see in a moment, this is in sharp contrast with fundamental laws.

Laws are patterns of a very special kind: they are systemic, not stray, and they do not change unless the things that satisfy or "obey" them are metamorphosed, i.e. change into things of a different kind. The laws of constructs are propositions, such as the associative and the commutative laws for the addition of numbers. On the other hand the laws of concrete or material things are properties of the latter—and not just any properties but their essential properties. These objective patterns, like the trends, are representable by propositions. We call them *law statements*. This difference, nonexistent in an idealist epistemology but central to a realistic one, is a particular case of the fact-construct difference. In every case a given factual item, be it state, event or pattern, is representable by alternative propositions, some truer than others.

A law statement is a hypothesis asserting that every thing of some kind has a certain property. The following are typical, though by no means exhaustive, forms of law statements found in science and technology. (*a*) "If x is a thing of kind K, and x possesses property P at any given time, then x conserves P at any other time". (*b*) "Every thing possessing property P possesses also property Q". (*c*) "All events of kind E have the same probability p". (*d*) "If x is a thing of kind A, and x undergoes a change of type P, x becomes a thing of kind B". (*e*) "The probability distribution of property P of things of kind K changes in the course of time according to such and such formula". (*f*) "Properties F and G of every thing of kind K are lawfully related to one another". (The mathematical form of this programmatic hypothesis is as follows. Let $f: A \to B$ and $g: A \to B$ be functions representing the properties F and G respectively. Then F and G are *lawfully related* if, and only if, there is a third function $h: A \to B$, representing another property of things of the same kind, such that g is the composition of h and f, i.e. $g = h \circ f$. See Figure 9.12.)

Law statements have occasionally been construed as prohibitions (Popper, 1959). Actually law statements tell us what is possible, from which we infer what is not. What happens is that, in certain popular versions, a handful of law statements are presented as statements to the effect that certain facts are impossible. Here are three well-known examples. (*a*) The

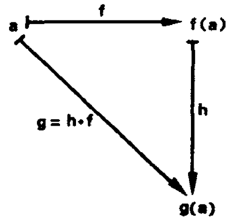


Fig. 9.12. Functions f and g represent properties of thing a. They are lawfully related if there is a third function h, which also represents a property of a, and such that its composition with f equals g.

principle of conservation of energy is sometimes translated into pragmatic terms (i.e. transformed into a nomopragmatic statement) for technological purposes, namely thus: "It is impossible to construct a perpetual motion machine of the first kind". But the original principle is affirmative, nonmodal, and refers to closed systems in general, not necessarily machines. (*b*) Pauli's exclusion principle states that the different components of a microsystem are in different states, or that the probability of their being in the same state is zero. But it can be reformulated in a more intuitive way, as "No two components of the same microsystem can be in the same state". (*c*) The competitive exclusion principle of ecology was originally formulated as "No two species can occupy exactly the same niche", or "Complete competitors cannot exist". A more correct (and more easily formalizable) formulation is: "Every species has its own peculiar niche" (or "There is a bijection between the set of species and the set of niches"). In short, law statements are not prohibitions but they can be used as premises for deriving prohibitions. By delimiting the set of lawfully possible events we *eo ipso* characterize the set of conceivable but factually impossible events.

The specific differences between a law statement and any other factual statement are generality, systemicity, and confirmation. More precisely, we adopt the following definition (Bunge, 1967a). A factual statement in a theory is a *law statement* if, and only if, (*a*) it is general in some respect (i.e. refers to every single member of a species, or extends over an entire region of spacetime), and (*b*) it has been satisfactorily confirmed (for the time being). New data may restrict the domain of truth of a law statement. They may show that the latter holds only between certain bounds of certain variables; or, on the contrary, they may show that the law statement holds beyond the domain originally postulated. And the data may also refute the

statement, in which case it is demoted from the rank of law statement, and the whole theory containing it is put in jeopardy.

This suffices to show how different laws are from definitions, with which they are sometimes confused. For example, Mach and his followers regarded Newton's second law of motion as a definition of "force". This is a mistake, for (*a*) there are a number of distinct force laws, which must be hypothesized separately from the law of motion, and (*b*) the latter, in conjunction with a definite force law, is empirically testable—and moreover it has been found false in the very small as well as in the very large. (For a criticism of Mach's attempt to define "force" and "mass" see Bunge (1966).)

The most precise representation of a law is with the help of functions. But not every function qualifies. Thus the function f that maps the set $\{a, b, c\}$ into the set $\{d, e\}$ in such a way that $f(a) = d$, $f(b) = e$, and $f(c) = d$, is a stray or *ad hoc* function that represents no intrinsic structure of its domain. We shall call a function *systemic* if it is the object of a theory and belongs to a system or family of functions, i.e. a set whose members are mutually related in precise ways—e.g. by recursion formulas. Otherwise the function will be said to be *stray*. For example, the trigonometric, spherical, cylindrical and hypergeometric functions form systems. The concept of a systemic function allows us to define lawfulness in this way: A set of factual items (states, events or processes) is *lawful* if it is representable by, or with the help of, systemic functions.

Probabilities are systemic for they are subject to the theory of probability. They occur in all *probabilistic* (or *stochastic*) law statements. Notwithstanding the opinion of distinguished philosophers (notably Nagel (1961) and Hempel (1962)) probabilistic law statements are universal in form and scope. For example a velocity distribution law states that, for *all* systems of a certain kind (e.g. molecules in a gas, or cars in dense traffic), the probability of an individual component having a velocity comprised between v and $v + \Delta v$ equals $p(v)$. Δv, where p is the probability distribution characteristic of the system. On the other hand statistical generalizations, which are empirical, are not universal. In fact, a typical statistical generalization is of the type "f percent of A's are B's". Every such generalization concerns a population as a whole, whereas the corresponding probabilistic laws concern every individual of a given population.

Randomness is not the opposite of lawfulness but one type of it. Thus every outcome of a coin flipping experiment has a definite probability. A sequence of events can be said to be *random* if every event in it has a definite

probability; otherwise the sequence is either chaotic or causal. (This is not the standard definition of randomness—but then there is no standard definition.) Randomness can exist objectively, as in the case of atomic and molecular events and of genic mutations; or it can be manufactured, as in the case of randomization and random sampling. Interestingly, random sampling can be used to uncover order, not only in factual science and technology but also in mathematics. This is the case with the Monte Carlo method for solving certain numerical problems. (*Example*: Suppose the problem is to compute the area of a region E of the plane included in the unit square. To do so select at random n points from the unit square, determine their positions x_i, and find out whether they belong to the given region E. That is, assign a value to the characteristic function χ_E of the set E to each sampled point. The Monte Carlo approximation is: Area of $E \cong (1/n) \sum_{i=1}^{n} \chi_E(x_i)$).

Laws, whether or not probabilistic, should not be mistaken for processes, in particular trajectories, life histories, and histories. Thus it is false that every one of us has his own laws: we all share a number of laws but differ in our circumstances, hence in our life histories. Every process can be analyzed into laws and circumstances. And every set of laws is compatible with many different processes, as many as different sets of circumstances. The particular circumstances are representable in the law statements by assigning special values to the parameters occurring in them or in their logical consequences. Thus there is presumably a single law of biological growth, but every pair $\langle$species, environment$\rangle$ is characterized by its own parameters. In other words, a single law covers an entire bundle of trajectories or histories. See Figure 9.13.

Given laws and data we can infer (tentatively) the corresponding processes. The inverse problem, to infer laws from descriptions of processes, is logically impossible because every law encompasses, in principle, infinitely many possible processes. In other words, law statements cannot be inferred from a knowledge of processes or histories. Hence they must be hypothesized—and checked. Far from being inferred, law statements are inference tools. For example, a recursion formula may allow one to infer the nth member of a family of functions from a knowledge of the $(n-1)$th or the $(n+1)$th or others. (Thus, the discrete analogue of the exponential function is a solution of the recursion formula $f_{n+1} = af_n$, where a is a parameter and n a natural number.) A law statement containing a time variable helps one predict the time at which the thing in

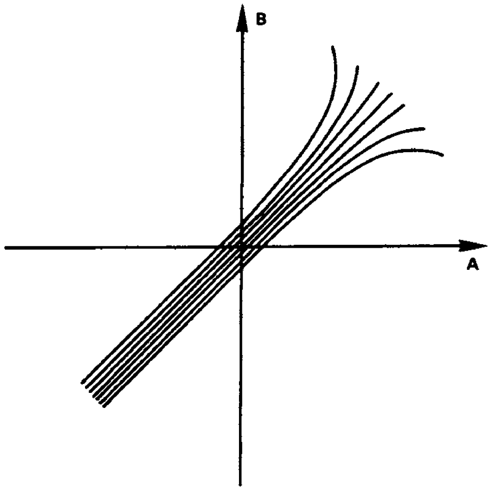


Fig. 9.13. A whole bundle of trajectories (e.g. life histories) in state space is described by a single law. Equivalently: a law is a bundle of possible histories.

question reaches a certain state. Likewise a law statement containing spatial coordinates helps us infer what happens at a given place from a knowledge of what is or was happening at another (e.g. on a surface surrounding the given place).

It is instructive to compare law statements with empirical generalizations, the former being the rationalist, the latter the empiricist paragon of scientific construct. Firstly, the empirical generalizations of the quantitative type, obtainable by interpolation techniques, are usually polynomials, which are stray not systemic functions. Not so the law statements, which contain systemic functions. Secondly, such polynomials contain adjustable parameters (one more than the degree of the polynomial). On the other hand the law statements, particularly of the mechanismic type, have a minimum of adjustable parameters. (The basic equations of electromagnetism and mechanics, whether classical or quantal, have no adjustable parameters at all: all the constants they contain are universal.) Thirdly, unlike the adjustable parameters in the empirical generalizations, the constants in the basic law statements have some factual meaning or other: they are universal constants. Fourthly, most basic law statements are mathematically more complicated than polynomials; even simple laws such as those of the oscillation period of a simple pendulum, and Snell's law of refraction, go far beyond polynomials.

Because of all these differences between law statements and empirical generalizations, the empiricist epistemology, which favors the latter and mistrusts or even rejects the former, does not fit the facts of scientific practice. Nor does critical rationalism, for which all hypotheses are groundless, none being better than any others except that some resist better the attempts at refuting them (Popper, 1959, 1963, 1974). As we saw before, a law statement is not just any hypothesis: it is a well-confirmed systemic hypothesis, hence one supported not only by data of a certain kind but also by all the data that support all the other hypotheses in the theory. Thus the laws of the hydrogen atom and the hydrogen molecule, of radiative decay and radioactive decay, of semiconductors and superconductors, derive from the basic laws of the quantum theory in conjunction with special assumptions; so, every time one of them is confirmed or refuted, its companions are confirmed or refuted.

Law statements can be classed in many ways: according to their logical form, mathematical properties, the kind of thing they refer to, their deductive power, generality, etc. (See Bunge (1963) for some such classings.) Here we shall mention only two distinctions, those between basic and derived, and between general and special. A *basic* law statement is one that occurs as a postulate or axiom in a general theory. Any logical consequence of basic law statements and definitions will be called a *derived* law statement. For example, the laws of conservation (of energy, linear momentum, etc.) follow from the equations of motion or field equations, which in turn follow from variational principles.

Our second partition of the set of law statements is into general and special. Since these terms are relative we must qualify them. We shall say that a law statement is *universal* in the collection F of factual items if it refers to every member of F—whether it holds for every one of them being another matter. The statement is *particular* in F if it refers to every member of a proper subset of F. The utmost general law statements are those concerning things of all kinds, from elementary particles to societies: they are the *universal* laws of ontology, such as "Every thing changes", "Nothing comes out of nothing", and "Every event is lawful". Then come the *generic* laws, i.e. those referring to all things of a given genus, e.g. bodies, electromagnetic fields, quantons, cells, societies, etc. Finally come the *specific* laws, i.e. those concerning all the members of a given species of thing, such as the so-called constitutive relations of continuum mechanics and electrodynamics, or all vertebrates, or all capitalist societies. Such specific law statements occur in all models or specific theories. Clearly, the

reference class of a specific law is included in that of the generic laws, which in turn are included in the reference class of the universal laws. And even the most specific of all laws, such as those concerning the motions of individual planets and their satellites, are general in some respect, e.g. they hold for all times. In other words, scientific laws, just like the conventions of social behavior, are not concerned with minutiae: *De minimis lex non curat*.

Our last type of pattern is the rule. Rules are precepts or recipes for doing something. Unlike law statements, rules do not describe, explain or forecast: they prescribe. (The difference between rules and factual laws can be elucidated with the help of the state space formalism used in Vols. 3 and 4. A *law* of things of kind K is a mapping from the state space S_K of the K's into itself, i.e. $L_K: S_K \to S_K$. Thus, if s is a state of a thing of kind K, i.e. s is in S_K, then "$L_K(s) = t$", where $t \in S_K$ represents the lawful passing of the thing from state s to state t. On the other hand a *rule* may be construed as a mapping from the set of all possible states of the system of interest to a set of possible actions of the agents dealing with it. More precisely, a rule prescribing the behavior of animals of kind M with respect to things of kind K is a mapping of the state space S_K into a collection A_M of actions in the behavioral repertoire of the individuals of kind M. I.e., $R_{KM}: S_K \to A_M$. Thus if s is in S_K, then "$R_{KM}(s) = a$", where $a \in A_M$ is the action prescribed by rule R_{KM} for individuals of kind M facing a system of kind K in state s. The distinction between rule and law disappears only when $M = K$ and $A_M = S_K$, i.e. when agents and patients are the same, which is the case of the rules of social action. We shall return to this important point in Vol. 6.)

An *instruction* can be construed as a set or a sequence of rules for handling problems of some kind. If the problem is conceptual, the instruction directs thinking; if practical, it controls doing. An *algorithm* is a special kind of instruction, namely a sequence of rules for performing effective operations on symbols, and sure to halt at the result that is being sought. For example, an instruction for extracting square roots, or for computing the values of some function, is an algorithm; a correct computer program too can be regarded as an algorithm. Algorithms are supposed to be consistent with laws, but they are not laws themselves; sometimes they are just clever law-abiding tricks. In particular, as usually understood the laws of mathematics are about concepts, whereas algorithms are rules for manipulating symbols designating concepts. Besides, one and the same set of laws is consistent with a variety of algorithms, some more efficient than others. Thirdly, whereas all the mathematical laws are parts of theories,

algorithms can be *ad hoc* and in any case they do not constitute hypothetico-deductive systems. So much for the algorithmic view of mathematics.

Animals and machines of certain kinds can be programmed or taught to follow instructions of certain types. Thus a Turing machine (an ideal computer) is characterized by its transition (or next state) function, which can be interpreted as embodying instructions of a certain kind—e.g. "When receiving input 1 while in state s, print 0 and jump to state t". A difference between an animal capable of learning and a Turing machine is that the latter has a fixed transition function built into it, whereas the animal can build and rebuild by itself a whole set of transition functions. (More on the difference between brain and machines in Vol. 4, Ch. 5.)

Whereas some rules are arbitrary, others are justifiable. The former are conventions, such as the rule that judges should wear black robes, and women discriminated against. A rule can be justified in either of two ways: practically (by success) or theoretically (by its compatibility with some theory). Thus until recently the medical rule "If you have a headache take an aspirin" was justified only practically. The theoretical justification of a rule consists in showing, with the help of some theory, *why* it is efficient. We call such rules *grounded*. And if the grounding theory is (sufficiently) true, the corresponding rules will be said to be *well grounded*; they may also be called *nomopragmatic* (instead of nomological) statements.

Take for example the so-called phase rule of thermodynamics: $C+2=P+F$, where C is the number of components, P the number of phases, and F the number of degrees of freedom of a system in equilibrium. Notwithstanding its name it is a law, for it belongs to a theory and concerns physical or chemical systems, not our actions on them. However, like many another law, the phase "rule" can be used to obtain certain desired effects. Thus, by increasing (decreasing) C we may cause an increase (decrease) in P. In other words, this law statement is the ground for a number of rules proper. All the rules employed in science and technology are supposed to be well grounded, for an empirical rule may be sheer superstition or, if effective, it may have unknown undesirable side effects. On the other hand the rules used or recommended by occultists, quacks, pseudoscientists, and unscientific ideologists are not well grounded.

The rule for obtaining a nomopragmatic statement from a nomological one is simple: Substitute a pragmatic expression for at least one of the expressions occurring in the law formula. Table 9.1 exhibits a few examples.

The relation between well grounded rules and the corresponding laws is

TABLE 9.1
Obtaining well grounded rules from law statements.

Law Schema	*Rule Schema*
1. When A rises B falls.	1. To decrease B, increase A.
2. If A happens B occurs.	2. When A is done B is found to happen.
3. For every x and some y, Fxy.	3. For any given x, some y may be found such that Fxy.
4. The rate of change of F equals G.	4. To vary the rate of change of F regulate G.
5. Every system has assembled from its components under suitable circumstances.	5. To manufacture a system bring its components together under suitable circumstances.

as follows. Consider an elementary rule schema of the form "To attain G use M", or G *per* M for short, where G is a goal and M a means. We stipulate that this rule is *well grounded* if, and only if, there is a law according to which M brings about G. In other words, the rule is well grounded only in case there is a reasonably well-confirmed theory containing a statement of the form "If M then G", where "M" and "G" are interpreted, in the theory, as properties, states, or events.

Note that the very same law statement that provides a basis for G *per* M supports also the dual rule *Non-G per non-M*, i.e. "To prevent G from being the case abstain from using M". (The contraposition of the law statement is not the ground of the dual rule. And in the latter the expressions of the form 'not-X' do not mean "anti-X" but merely the absence of X.) For example, "The output of an economic system is an increasing function of both capital and labor" provides the ground for two rules: "To increase output increase capital or labor", and "To decrease output decrease capital or labor". This practical ambivalence of law statements cannot be accounted for by pragmatism, which mistakes laws for prescriptions.

To say that a rule is *based* or *grounded* on a law statement is not to say that it follows from the latter—if only because every law statement supports a pair of rules. This point is of interest not only to the philosophy of technology but also to all the normative disciplines. In particular, to demand that the ethical and legal norms be well grounded—e.g. on psychology and sociology—does not amount to requiring that they be deducible from them. In other words, even if we wish the normative

disciplines to be based on science we cannot reduce them to the latter. (This point will be discussed in detail in Vol. 7.)

Laws and rules may combine in yet another way, namely as rules helping or blocking the formulation of hypotheses expected to attain the rank of law. One such blocking rule is the commandment that scientific formulas ought to involve only functions representing observable properties. (Unfortunately for this rule all the basic concepts of the fundamental physical theories, in particular those of field intensity, Lagrangian, and state function, represent unobservables.) Another such rule, but this time a fertile one, is the correspondence principle: "Every new theory should contain, in some limit or other, the true, or approximately true, formulas of its predecessors". A third is the general covariance principle: "Every basic law ought to be frame-independent, i.e. to remain invariant under changes in frame of reference". This rule instructs us to get rid of constants specifying particular (e.g. egocentric) facts, and it has been a powerful stimulus to the construction of coordinate-free geometries. A fourth is the rule of equipresence, widely used in contemporary thermomechanics: "A magnitude present as an independent variable in one constitutive relation should be so present in all". This rule suggests including in every proposed constitutive relation all the variables occurring in other formulas of the same area.

Neither of the above principles represents a law of nature: all of them are heuristic rules concerning scientific hypotheses and theories. We call them *metanomological* statements. Clearly, they are rules or norms, hence efficient, inefficient, or counterproductive but not true or false. They are rules to be justified exclusively by their success in suggesting or eliminating hypotheses aiming at capturing objective patterns: they cannot be put to the test in the laboratory or in the field. On the other hand the statement that the basic equations of classical mechanics remain invariant under a transformation of coordinates mirroring an exchange of inertial (Galilean) frames of reference, is a metanomological statement that can be proved with pencil and paper. The same holds for the statement of the Lorentz invariance of the Maxwell equations, and of the charge-parity-time reversal invariance of all the basic local field theories. These metatheoretical statements are true, and their truth can be checked by purely conceptual operations, so they do not represent any real patterns. So much so that one can build any number of theories abiding by such principles and having nothing to do with reality. (For more on metanomological statements see Bunge (1963, 1967a).)

This concludes our treatment of laws and rules, subjects that, despite their centrality, are neglected by most philosophers—or, worse, they are mixed up with extraneous topics such as induction, counterfactuals, and possible worlds.

4. Concluding remarks

Most of us go through life without ever encountering well-constructed classifications or hypothetico-deductive systems. We get by with vague classifications that are neither exclusive nor exhaustive, and with stray conjectures or tangled bodies of propositions that we dignify with the label 'theory'. In Antiquity only a handful of thinkers, such as Aristotle, Euclid and Archimedes, proposed reasonable classifications or well-built theories. In the Middle Ages only a few theologians, in the first place Thomas Aquinas, constructed hypothetico-deductive theories—albeit speculative ones. And in the Modern period there was no general theory before Newton's, and no general and correct classification before Linnaeus'. But from then on classifications and theories flourished and multiplied. In particular, theory has become so central to science and technology, that any discipline that lacks theories is properly regarded as nonscientific or, at best, protoscientific.

However, the careful and systematic analysis of classifications and theories did not begin until the 19th century, in the first place by Whewell (1847). Before him philosophers had no adequate conception of theories. Thus Bacon cared for none and explicitly rejected the few scientific theories of his time, and the British empiricists ignored factual theories altogether. Descartes and Leibniz thought scientific theories were a priori mathematical structures, and never thought it necessary to subject them to experimental tests. Kant did conceive of scientific theory as a synthesis of reason and experience—but the wrong kind of synthesis. Indeed he thought theories can account for appearance not reality, and he failed to understand the scientific theories of his time for want of mathematical knowledge.

We are better placed than our ancestors to analyze the logical structure of classifications and theories, because in the meantime logicians and mathematicians have forged the necessary tools of analysis, such as the predicate calculus, the theories of sets, lattices, filters, and ideals, model theory, and the calculus of theories. (See Tarski, 1956; Rasiowa and Sikorski, 1963; Bell and Slomson, 1969.) Yet, all this concerns structure not

content. Accordingly only logicians and mathematicians have attained an adequate (though surely not final) understanding of their own systematizations. Many scientists, technologists and philosophers are often still confused or wrong about the nature of scientific and technological systematizations. This is partly explainable by the intrinsic difficulty of the subject: not only the formal structure or formalism but also the content is to be accounted for. (Hilbert said that he found physics much more difficult than mathematics.) But there is also the weight of tradition: mathematics has always been helped by rationalism and never been seriously hampered by empiricism. On the other hand factual science, technology and philosophy are still being hindered by empiricism.

The misunderstanding and even mistrust of theory among many philosophers is apparent in the half-century old debate on the status of theoretical concepts. Many are still worried by the fact that theories happen to contain theoretical concepts. The less sophisticated among them get away by decreeing that there is no problem: that every scientific theory concerns observations or measurements, not autonomously existing things (e.g. Simon and Groen, 1973). But of course this operationist reform (or rather destruction) does not work, because (*a*) measurement instruments are designed and operated with the help of theories, (*b*) we often find it necessary to correct measurement results in the light of theories, and (*c*) the basic concepts of the most advanced theories happen not to represent observable things or properties. (See e.g. Bunge, 1973a.)

Other philosophers recognize the existence of theoretical concepts but try to eliminate them by various ingenious techniques, of which the following are the best known. (*a*) Define the theoretical concepts in terms of observational ones, the way Mach tried to define "mass" as a ratio of accelerations. (*b*) Construct Ramsey sentences with the help of the empirical concepts of the theory (e.g. Carnap, 1966). (*c*) Pretend that the theory is composed of formulas that are either observational or theoretical, and apply Craig's theorem (Craig, 1956). None of these techniques work. Mach's because it presupposes the very law statements used to "define" the suspect concepts. (See Bunge, 1966). Ramsey's because the theoretical concepts must have been introduced, either as primitives or as defined concepts, before the postulates of the theory can be stated and conjoined to form a Ramsey sentence. And Craig's technique fails because it rests on the false assumption that it is possible to partition the set of formulas of a theory into observational and theoretical, while actually no factual theories contain strictly observational formulas. And all three

TABLE 9.2
Characterization of epistemologies with respect to their theses on the representability of scientific systematizations.

Epistemology	*Thesis*
Conventionalism	Do not represent: just colligate data.
Empiricism	Represent human experiences.
Pragmatism	Do not represent: are just tools for action.
Naive realism	Represent directly and faithfully.
Critical realism	Represent indirectly and, at best, approximately.

theory demolition techniques fail because they ignore that factual theories happen to be about concrete things existing whether we know it or not, and having properties that can be represented only by sophisticated concepts. (For further criticisms see Bunge (1967a, 1973a, 1973b), and Tuomela (1973).) This view, critical realism, contrasts with the alternative views as shown in Table 9.2.

Finally let us define the reference, extension, scope, and power of a factual theory, and let us evaluate theorizing in general. The *reference* (class) of a theory is of course the collection of its referents. These are not only the actual and investigated ones but all of the possible things of a kind—e.g. all the possible biopopulations in the case of the genetics of populations. (See Vol. 1, Ch. 2.) On the other hand the *extension* of a theory T may be defined as the collection of referents of T for which T has been confirmed reasonably well, i.e. for which T holds with good approximation. Note that the extension of a theory is a subset of its reference, and that it is a variable collection rather than a fixed set, for the incoming empirical information is likely to expand or shrink, in the course of time, the range of things that are adequately covered by the theory.

The *scope* of a theory may be defined as the set of events it allows or regards as lawful, i.e. the event space it assigns to its referents. (The event space of things of kind K is properly included in the cartesian product of the state space S_K by itself—i.e. it is the set of pairs of states compatible with the laws of things of kind K. See Vol. 3, Ch. 5.) The more general a theory, the greater its scope. As a theory is adjoined subsidiary hypotheses accounting for details of its referents, its scope shrinks: it tells us more and more about less and less. We need theories of all scopes except the two extremes of complete restriction ("Nothing ever happens") and total permissiveness ("Anything goes"). The former extreme must be avoided because it leaves

nothing to be investigated, the latter because it has no explanatory power and consecrates magic.

The *power* of a theory is measurable by the number of bits of knowledge it systematizes—in particular the number of precursor hypotheses it puts together—and the number of new items it introduces. Of all available scientific theories the quantum theory and the synthetic theory of evolution are the most powerful. But such power, though great, is not unlimited: no theory can embrace all the knowledge in a given field. The reason for this limitation is that a factual theory should contain no data. All we demand of theories is that they contain blanks that can be filled out with empirical data every time we apply them to solve particular problems. (Recall Section 2.2.)

As for the value of theories, different people value them differently. To theoretical scientists and rationalist philosophers, theory is at the very top, for it constitutes the best possible systematization, extrapolation, and exactification of a body of knowledge. To descriptive scientists, empiricists, philosophers, and laymen, theory is a matter of either indifference or embarrassment for overreaching experience, and so it is at the very bottom of their scale of values. And to workers in experimental science and technology, theory is right in the middle, for they are guided by theories and they know their discoveries or inventions motivate the building and correction of theories. The realist epistemologist values equally theory and experiment: to him it is a truism that theory without empirical support is sheer speculation, whereas empirical operation without theoretical backing is a waste of time.

This concludes the first part of our study of epistemology and methodology. The second part, titled *Understanding the World*, examines the ways we account for facts and check and evaluate our hypotheses, theories and proposals. It also discusses the varieties of human knowledge and of changes of it. Finally, it formulates and defends the general view resulting from our study, namely scientific realism.

APPENDIX 1

THE POWER OF MATHEMATICS IN THEORY CONSTRUCTION: A SIMPLE MODEL OF EVOLUTION

To illustrate the power of mathematics in the construction of theories consider the following simple-minded model of evolution representing the two main evolutionary "forces", namely random mutation and selection by the environment. (This model is quite different from the one in Vol. 4, Ch. 3, Section 3.4.) The evolving units may be populations of molecules, stars, organisms, or social systems: the theory does not specify the precise mutation or selection mechanisms. We assume that the components of such a population are subject to mutation, whether genic or of some other type, spontaneous or induced by environmental agents, but in any case "blind" or random, in the sense that the probability of a mutation is independent of its survival value for the component. We also suppose, for the sake of simplicity, that there is a single kind of mutation and that some of the mutants happen to be viable, so that in the course of time a second population emerges, namely that of mutants. Call the latter M and that of normals N. Suppose also that the mutations are irreversible and that the mutants themselves are stable, i.e. immune to any further mutations, so that they contribute neither to the normal population nor to any third group. Besides, assume that all the normals have the same probability p of originating mutating offspring, and that on the average each normal has a descendants, where $a > 1$. Finally, suppose that all of the viable individuals endure, regardless of their age, for a large number g of generations. Note that our model contains no less than six simplifying assumptions, most of which are at best only approximately true.

Our problems are (*a*) to trace the speciation process, i.e. the growth of the M population, and (*b*) to estimate the mutation parameter p from a count of the fraction of mutants in the total population after g generations. We start by noting that, since the probability of a mutation from one generation to the next is p, the probability that no such change occurs during the same period is $1-p$, and the probability that a normal individual will originate normal offspring at generation n will be

$$v_n = (1-p)^n.$$

Consequently the probability that an organism taken at random from the total population at generation n be a mutant, is the complement of v_n to

unity, i.e.

$$\mu_m = 1 - (1-p)^n.$$

Recall now that every normal individual produces, on the average, a descendants. Hence the average number of normals at generation n is

$$N_n = a^n \nu_n = a^n(1-p)^n.$$

Therefore the total average of normals for all g generations is

$$N = \sum_0^g N_n = \sum_0^g a^n(1-p)^n = \frac{(a(1-p))^{g+1}-1}{a(1-p)-1}.$$

To compute the corresponding number M of mutants produced at generation g assume that on the average every mutant produces b offspring, where b is in general different from a—in fact greater than a if the mutants are fitter than the normals. Hence the average number of mutants at generation n is

$$M_n = b^n \mu_n = b^n(1-(1-p)^n).$$

Consequently the average number of mutants at generation g is

$$M = \sum_0^g M_n = \sum_0^g b^n(1-(1-p)^n) = \sum_0^g b^n - \sum_0^g (b(1-p))^n$$
$$= \frac{b^{g+1}-1}{b-1} - \frac{(b(1-p))^{g+1}-1}{b(1-p)-1}.$$

What can be observed directly are the totals N and M. From them we can estimate the mutation parameter p as follows. To simplify the calculations make the unrealistic pretense that the mutants are just as fit as the normals, i.e. set $a = b$, and use the previous results to form the fraction of normals:

$$\frac{N}{N+M} = \frac{\dfrac{(a(1-p))^{g+1}-1}{a(1-p)-1}}{\dfrac{a^{g+1}-1}{a-1}} \cong (1-p)^{g+1},$$

since, for small p, $a(1-p) \cong a$. Finally we obtain

$$p \cong 1 - \left(\frac{N}{N+M}\right)^{1/(g+1)}.$$

For example, for $p = 0.001$ and $g = 99$, the total fraction of mutants proves

to be 0.1, an appreciable number. If the mutation probability is tenfold, the mutant population attains over one third of the total. But even a small mutation probability can be compensated for by a large fitness or fertility value. In other words, a favorable environment can make up for a small chance of an individual change.

Note the following points. First, the entire model involves only a couple of theorems in the elementary calculus of probability and a handful of algebraic manipulations. Second, every assumption is translated into mathematical terms, and every mathematical result is translated back into scientific terms, i.e. is assigned a factual interpretation. Third, the last two formulas, which relate the unobserved probability p to the observable frequency $N/(N+M)$, is characteristic of this theory. Different stochastic theories contain different relations between the two quantities—or none. (Therefore, it is hopeless to try and define probability as long term frequency, the way empiricists have tried. In general, theoretical concepts cannot be defined in terms of empirical ones or conversely. But they can be related.)

THE PROSE IDENTIFYING THE VARIABLES

Take the hypothesis that frustration builds up latent aggression (i.e. aggressiveness). This hypothesis may be taken as the precursor of a theory composed of mathematical formulas and semantic formulas. Let us proceed to build the simplest possible such theory.

We start by defining frustration as the percentage of unfulfilled desired acts, and postulate that latent aggression (or aggressive disposition) is proportional to manifest aggression. That is, we set

$$F = (D - P)/D \qquad \textit{Definition} \quad (1)$$

and

$$A = aN, \qquad \textit{Hypothesis} \quad (2)$$

where "D" and "P" stand for desired and performed acts of some kind respectively, while "N" stands for the total number of aggressions and "a" their intensity, assumed to be constant for the sake of simplicity.

We can now state the original (precursor) hypothesis in a definite fashion:

$$A = bF, \qquad \textit{Hypothesis} \quad (3)$$

where b is a positive real number. Obviously, this assumption of mere proportionality between aggressiveness and frustration is too simple to be true. But our purpose is semantic insight not factual truth.

Substituting (1) and (3) into (2) we obtain a testable consequence:

$$N = b(D - P)/aD. \qquad \textit{Theorem} \quad (4)$$

Finally we may collect the above formulas, restate them more carefully, and surround them with the requisite mathematical and semantical formulas. The result is this:

Primitive concepts
Set S of subjects
Number D of desired acts of a given kind per unit time
Number P of desired and performed acts of a given kind per unit time
Number N of aggressive acts per unit time

Defined concept

Frustration $F = (D - P)/D$

Postulates

Mathematical 1 $D, P: S \to \mathbb{N}$.
Mathematical 2 $A, a: S \to \mathbb{R}^+$
Mathematical 3 $b \in \mathbb{R}^+$
Factual 1 For every x in S, $F(x) = a(x)N(x)$.
Factual 2 For every x in S, $A(x) = bF(x)$.
Semantic 1 "$A(x)$" represents the aggressiveness of subject x.
Semantic 2 "$F(x)$" represents the frustration felt by subject x.
Semantic 3 "$a(x)$" represents the intensity of the aggressive acts committed by subject x.

Theorem

For every x in S, $N(x) = b[D(x) - P(x)]/a(x)D(x)$.

The same mathematical and factual axioms may occur in totally different contexts accompanied by corresponding different semantic axioms. And if the former are changed then the latter too may have to be modified. For example, instead of construing frustration and aggressiveness as numerical variables we may work with the sets of aggressive acts and desired acts, whether performed or frustrated. And we may then try a probabilistic version of the precursor hypothesis, such as, e.g.,

For any subject x, every a in A, every f in F, and every p in P,

$$Pr(a|f) = a(x)[1 - Pr(a|p)], \quad \text{with } a(x) > 1.$$

The semantic assumptions are now

*S*1 "$Pr(a|f)$" represents the propensity (tendency) of a subject to commit an aggressive act when he feels a frustrated desire f.
*S*2 "$Pr(a|p)$" represents the propensity (tendency) of a subject to commit an aggressive act when he has fulfilled desire p.
*S*3 "$a(x)$" represents the natural aggressiveness of the subject x.

The point is that every formalizations calls for its own set of semantic formulas, without which we would not know what it represents.

BIBLIOGRAPHY

Abelson, P. H. (1965). Relations of group activity to creativity in science. *Proc. Amer. Acad. Arts and Science* **94**: 603–614.

Ackoff, R. L. (1974). *Redesigning the Future: A Systems Approach to Societal Problems*. New York: Wiley-Interscience.

Ackoff, R. L. (1978). *The Art of Problem Solving*. New York: John Wiley and Sons.

Adams. J. L. (1980). *Conceptual Blockbusting*, 2nd ed. New York: W. W. Norton and Co.

Agassi, J. (1975). *Science in Flux*. Dordrecht and Boston: Reidel.

Alcock, J. (1981). *Parapsychology: Science or Magic?* Oxford: Pergamon Press.

Amari, S.-I., and A. Takeuchi (1978). Mathematical theory on formation of category detecting nerve cells. *Biol. Cybernetics* **29**: 127–136.

Aqvist, L. (1965). A New Approach to the Logical Theory of Interrogatives. Mimeo. University of Uppsala.

Asch, S. E. (1952). *Social Psychology*. New York: Prentice-Hall.

Aubrey, J. (1949). *Aubrey's Brief Lives*. O. Lawson Dick, Ed., London: Secker and Warburg.

Bandura, A. (1974). Behavior theory and the models of man. *Amer. Psychol.* **29**: 859–869.

Baranyi, A., and O. Fehér (1981). Synaptic facilitation requires paired activation of convergent pathways in the neocortex. *Nature* **290**: 413–415.

Barber, T. X. (1978). Hypnosis, suggestions, and psychosomatic phenomena: a new look from the standpoint of experimental studies. *Am. J. Clinical Hypnosis* **21**: 13–27.

Barber, T. X., and S. C. Wilson (1979). Guided imaging and hypnosis. In A. A. Sheikh and J. T. Shaffer, Eds., *The Potential of Fantasy and Imagination*. New York: Brandon House.

Barnes, B. (1982). *T. S. Kuhn and Social Science*. New York: Columbia University Press.

Bartlett, F. C. (1932). *Remembering*. Cambridge: Cambridge University Press.

Bartlett, F. (1958). *Thinking*. New York: Basic Books.

Battro, A. M., S. P. Netto, and R. J. A. Rozestraten (1976). Riemannian geometries of variable curvature in visual space. *Perception* **5**: 9–23.

Battro, A. M. (1979). Psicología, geometría y filosofía del espacio visual. *Revista Latinoamericana de Filosofía* v: 19–31.

Bechtereva, N. P. (1978). *The Neurophysiological Aspects of Human Mental Activity*. New York: Oxford University Press.

Becker, G. (1964). *The Human Capital*, 2nd ed. New York: Columbia University Press.

Beer, C. G. (1973). A view of birds. In A. Pick, Ed., *Minnesota Symposia on Child Psychology*, Vol. 7, pp. 47–86. Minneapolis: University of Minnesota Press.

Békesy, G. von (1967). *Sensory Inhibition*. Princeton: Princeton University Press.

Békesy, G. von (1968). Problems relating psychological and electrophysiological observations in sensory perception. *Perspectives in Biology and Medicine* **11**: 179–194.

Bell, J. L., and A. B. Slomson (1969). *Models and Ultraproducts*. Amsterdam: North-Holland.

Bellugi, U., and E. S. Klima (1979). Language: Perspectives from another modality. In *Brain and Mind*, Ciba Foundation Symposium 69, pp. 99–117.

Berger, P. L., and T. Luckmann (1966). *The Social Construction of Reality*. New York: Doubleday.

Berlyne, D. E. (1954). A Theory of Human Curiosity. *Brit, J. Psychol.* **45**: 180–191.

Bertalanffy, L. von (1955). An essay on the evolution of categories. *Phil. Science* **22**: 243–263.

Bindra, D. (1974). A motivational view of learning, performance, and behavior modification. *Psychol. Rev.* **81**: 199–213.

Bindra, D. (1976). *A Theory of Intelligent Behavior*. New York: Wiley Interscience.

Bindra, D., Ed. (1980). *The Brain's Mind: A Neuroscience Perspective on the Mind–Body Problem*. New York: Gardner Press.

Bindra, D. (1981). Ape language. *Science* **211**: 86.

Bindra, D. (1982). Cognitivism: Its origin and future in psychology. *Annals of Theoretical Psychology*, forthcoming.

Birch, H. G. (1945). The relation of previous experience to insightful problem-solving. *J. Comparative Psychology* **38**: 367–383.

Birch, H. G., and H. S. Rabinowitz, (1951). The negative effect of previous experience on productive thinking. *J. Experimental Psychology* **41**: 121–125.

Black, M. (1962). *Models and Metaphors*. Ithaca, New York: Cornell University Press.

Bliss, T. V. P. (1979). Synaptic plasticity in the hippocampus. *Trends in Neurosciences* **2**: 42–45.

Bloor, D. (1977). The regulatory function of language. In Morton and Marshall, Eds. (1977), pp. 73–98.

Bohr, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **48**: 686–702.

Boltzmann, L. (1905). *Populäre Schriften*. Leipzig: Barth.

Bolzano, B. (1837). *Wissenschaftslehre*, 4 vols. Repr.: Lepzig, Meiner, 1929.

Botha, R. P., and W. K. Winkler (1973). *The Justification of Linguistic Hypotheses*. The Hague: Mouton.

Botkin, J. W., M. Elmandjra, and M. Malitza (1979). *No Limits to Learning*. Oxford: Pergamon Press.

Bouchard, T. J., Jr., and M. McGue (1981). Familial studies of intelligence: a review: *Science* **212**: 1055–1059.

Boudon, R. (1967). *L'analyse mathèmatique des faits sociaux*. Paris: Librairie Plon.

Brewer, D. (1980). On the theory and practice of innovation. *Technology in Society* **2**: 337–363.

Bridgman, P. W. (1927). *The Logic of Modern Physics*. New York: Macmillan.

Brines, M. L., and J. L. Gould (1979). Bees have rules. *Science* **206**: 571–573.

Brown, H. I. (1977). *Perception, Theory and Commitment. The New Philosophy of Science*. Chicago: Precedent Publ. Inc.

Bruner, J. S., J. J. Goodnow, and G. Austin (1959). *A Study of thinking*. New York: John Wiley and Sons.

Bunge, M. (1959a). *Causality*. Cambridge, Mass.: Harvard University Press. Rev. ed.: *Causality in Modern Science*. New York: Dover, 1979.

Bunge, M. (1959b). *Metascientific Queries*. Evanston, Ill.: Charles C. Thomas.

Bunge, M. (1962a). *Intuition and Science*. Englewood Cliffs, N.J.: Prentice-Hall. Repr.: Westport, Conn.; Greenwood Press, 1975.

Bunge, M. (1964). Phenomenological theories. In M. Bunge, Ed., *The Critical Approach*, pp. 234–254. New York: The Free Press.

Bunge, M. (1966). Mach's critique of Newtonian mechanics. *American Journal of Physics* **34**: 585.

Bunge, M. (1967a). *Scientific Research I. The Search for System*. New York: Springer-Verlag.

Bunge, M. (1967b). *Scientific Research II. The Search for Truth*. New York: Springer-Verlag.

Bunge, M. (1967c). *Foundations of Physics*. New York: Springer-Verlag.

Bunge, M. (1967d). The structure and content of a physical theory. In M. Bunge, Ed., *Delaware Seminar in the Foundations of Physics*, pp. 15–27. New York: Springer-Verlag.

Bunge, M. (1968). Scientific laws and rules. In R. Klibansky, Ed., *Contemporary Philosophy* II, pp. 128–140. Firenze: La Nuova Italia Editrice.

Bunge, M. (1970). Virtual processes and virtual particles: real or fictitious? *Intern. J. Theoretical Physics* **3**: 507–508.

Bunge, M. (1971). A mathematical theory of dimensions and units. In M. Bunge, Ed., *Problems in the Foundations of Physics*, pp. 1–16. New York: Springer-Verlag.

Bunge, M. (1973a). *Philosophy of Physics*. Dordrecht: Reidel.

Bunge, M. (1973b). *Method, Model and Matter*. Dordrecht: Reidel.

Bunge, M., Ed. (1973d). *Exact Philosophy*. Dordrecht: Reidel.

Bunge, M. (1973e). Review of Suppes' (1970). *Brit. J. Phil. Sci.* **24**: 409–410.

Bunge, M. (1974a). *Sense and Reference* (Treatise, Vol. 1). Dordrecht: Reidel.

Bunge, M. (1974b). *Interpretation and Truth* (Treatise, Vol. 2). Dordrecht: Reidel.

Bunge, M. (1977a). *The Furniture of the World* (Treatise, Vol. 3). Dordrecht and Boston: Reidel.

Bunge, M. (1977b). Levels and reduction. *Am. J. Physiol.* **233**: R75–82.

Bunge, M. (1977d). The philosophical richness of technology. In F. Suppe and P. D. Asquith, Eds., *PSA 1976*, Vol. 2, pp. 153–172. East Lansing, Mich.: Philosophy of Science Assn.

Bunge, M. (1977f). General systems and holism. *General Systems* **XXII**: 87–90.

Bunge, M. (1978b). A model of evolution. *App. Math. Modell.* **2**: 201–204.

Bunge, M. (1980a). *The Mind-Body Problem*. Oxford and New York: Pergamon Press.

Bunge, M. (1980b). *Ciencia y desarrollo*. Buenos Aires: Siglo Veinte.

Bunge, M. (1981a). *Scientific Materialism*. Dordrecht and Boston: Reidel.

Bunge, M. (1981b). Analogy between systems. *Intern. J. General Systems* **7**: 221–223.

Bunge, M. (1981c). Four concepts of probability. *Appl. Math. Modell.* **5**: 306–312.

Bunge, M. (1981e). Review of Fleck (1935). *Behav. Sci.* **26**: 178–180.

Bunge, M. (1982a). The revival of causality. In G. Fløistad, Ed., *Contemporary Philosophy*, Vol. 2, pp. 135–155. The Hague: Martinus Nijhoff.

Bunge, M. (1982b). *Economía y filosofía*. Madrid: Tecnos.

Bunge, M. (1983) *Lingüística y filosofía*. Barcelona: Ariel.

Calvin, W. H. and G. A. Ojemann (1980). *Inside the Brain*. New York: New American Library.

Campbell, D. T. (1959). Methodological suggestions from a comparative psychology of knowledge processes. *Inquiry* **2**: 152–182.

Campbell, D. T. (1974). Evolutionary epistemology. In Schilpp, Ed., pp. 413–463.

Campbell, D. T. (1977). *The William James Lectures*, Harvard University. Preprint courtesy of Professor Campbell.

Campbell, D. T. (1979). A tribal model of the social system vehicle carrying scientific knowledge. *Knowledge* **1**: 181–217.

Čapek. M. (1968). Ernst Mach's biological theory of knowledge. *Synthese* **18**: 171–191.

Carnap, R. (1966). *Philosophical Foundations of Physics*. M. Gardner. Ed. New York: Basic Books.

Carterette, E. C., and M. P. Friedman, Eds. (1976). *Handbook of Perception*, Vol. VII. New York: Academic Press.

Causey, R. L. (1977). *Unity of Science*. Dordrecht-Boston: Reidel.

Chapanis, N. P., and A. Chapanis (1954). Cognitive dissonance: five years later. *Psychol Bull.* **61**: 1–22.

Chapman, L. J., and J. P. Chapman (1959). Atmosphere effect re-examined. *J. Exp. Psychol.* **58**: 220–226.

Chomsky, N. (1975). *Reflections on Language*. New York: Pantheon Books.

Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.

Claparède, E. (1934). La genèse de l'hypothèse. Étude experimentale. *Archives de psychologie* **24**: 1–155.

Clark, E. V. (1977). First language acquisition. In Morton and Marshall, Eds., pp. 1–72.

Cohen, N. L., and L. R. Squire (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that. *Science* **210**: 207–209.

Cole, S., J. R. Cole, and G. A. S. Simon (1981). Chance and consensus in peer review, *Science* **214**: 881–886.

Coleman, J. S. (1980). The structure of society and the nature of social research. *Knowledge* **1**: 333–350.

Conel, J. L. (1939–1967). *The Postnatal Development of the Human Cerebral Cortex*, 8 volumes. Cambridge, Mass.: Harvard University Press.

Craig, W. (1956). Replacement of auxiliary expressions. *Philosophical Review* **65**: 38–55.

Currie, G. (1978). Popper's evolutionary epistemology. *Synthese* **37**: 413–431.

Davenport, L. D. (1979). Superstitious bar pressing in hippocampal and septal rats. *Science* **205**: 721–723.

Davidson, D. (1963). Actions, reasons, and causes. *J. Phil.* **60**: 685–700.

Dawson, M. E., and J. J. Furedy (1976). The role of awareness in human differential autonomic classical conditioning: the necessary-gate hypothesis. *Psychophysiology* **13**: 50–53.

Denenberg, V. (1979). Paradigms and paradoxes in the study of behavioral development. In E. Thoman, Ed., *Origins of the Infant's Social Responsiveness*, pp. 251–289. Hillsdale, N.J.: L. Erlbaum Assoc.

Denison, E. F. (1962). *The Sources of Economic Growth in the United States*. New York: Committee for Economic Development.

Dickinson, A. (1980). *Contemporary Animal Learning Theory*. Cambridge: Cambridge University Press.

Dickinson, A., and N. J. Mackintosh (1978). Classical conditioning in animals. *Annual Rev. of Psychology* **29**: 587–612.

Dodwell, P. C. (1975). Contemporary theoretical problems in seeing. *Handbook of Perception*, Vol. V: 57–77. New York: Academic Press.

Dreistadt, R. (1969). The use of analogies and incubation in obtaining insights in creative problem solving. *Journal of Psychology* **71**: 159–175.

Duhem, P. (1914). *Essai sur la notion de théorie physique*, 2nd ed. Paris: Hermann.

Duncker, K. (1945). On problem-solving. *Psychological Monograph* 58, No. 270.

Dyson, F. J. (1958). Innovation in physics. *Scientific American* **199**: 3.

Eberle, R. D. Kaplan, and R. Montague (1961). Hempel and Oppenheim on explanation. *Phil. Sci.* **28**: 418–428.

Eccles, J. C. (1978). *The Human Mystery*. New York: Springer-Verlag.

Eccles, J. C. (1980). *The Human Psyche*. Berlin, Heidelberg, New York: Springer-Verlag.

Einhorn, H. J., and R. M. Hogarth (1978). Confidence in judgment: persistence in the illusion of validity. *Psychol. Rev.* **85**: 395–416.

Ellis, A. E. (1963). An introduction to the principles of scientific psychoanalysis. In S. Rachman, Ed., *Critical Essays on Psychoanalysis*: Oxford: Pergamon Press.

Evans, E. F., and I. C. Whitfield (1964). Classification of unit responses in the auditory cortex of the unanesthesized and unrestrained cat. *J. Physiol.* **171**: 476–493.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd. ed. New York: Wiley.

Ferrater Mora, J. (1967). *El ser y el sentido*. Madrid: Revista de Occidente.

Festinger, L. (1957). *Cognitive Dissonance*. Stanford: Stanford University Press.

Feyerabend, P. K. (1975). *Against Method*. Repr., London: Verso, 1978.

Feyerabend, P. K. (1981). *Philosophical Papers*, 2 Vols., Cambridge: Cambridge University Press.

Finetti, B. de (1972). *Probability, Induction and Statistics*. New York: John Wiley.

Fleck, L. (1935). *Genesis and development of a Scientific Fact*. Transl. F. Bradley and T. J. Trenn. Chicago: University of Chicago Press, 1979.

Flohr, H., and W. Precht, Eds. (1981). *Lesion-Induced Neuronal Plasticity in Sensorimotor Systems*. New York: Springer-Verlag.

Fodor, J. (1975). *The Language of Thought*. New York: Thomas Y. Crowell.

Fodor, J. (1981). The mind–body problem. *Scientific American* **244, 1**: 114–123.

Garcia, J. (1981). Tilting at the paper mills of Academe. *American Psychologist* **36**: 149–158.

Gardner, R. A., and B. T. Gardner (1969). Teaching sign language to a chimpanzee. *Science* **165**: 664–672.

Ghiselin, M. T. (1981). Categories, Life, and thinking. *Behav. and Brain Sci.* **4**: 269–283.

Gibson, J. (1966). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

Gibson, J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gilbert, C. D., and T. Wiesel (1979). Morphology and intracortical projections of functionally characterised neurones in the cat visual cortex. *Nature* **280**: 120–125.

Glass, L. (1975). Classification of biological networks by their qualitative dynamics. *J. Theoretical Biology* **54**: 85–107.

Goddard, G. V. (1980). Component properties of the memory machine: Hebb revisited. In Jusczyk and Klein, Eds., pp. 231–247.

Goody, J. (1977). *The Domestication of the Savage Mind*. Cambridge: Cambridge University Press.

Gould, S. J. (1978). Morton's ranking of races of cranial capacity. *Science* **200**: 503–509.

Greenberg, J. H., M. Reivich, A. Alavi, P. Hand, A. Rosenquist, W. Rintelmann, A. Stein, R. Tusa, R. Dann, D. Christman, J. Fowler, B. McGregor, and A. Wolf (1981). Metabolic mapping of functional activity in human subjects, etc. *Science* **212**: 678–680.

Greenwell, J. R. (1980). Academia and the occult: An experience at Arizona. *The Skeptical Inquirer* **5**: 39–45.

Gregory, R. (1970). *The Intelligent Eye*. New York: MacGraw-Hill.

Gruber, H. E. (1981). On the relation between "aha experiences" and the construction of ideas. *History of Science* **19**: 41–59.

Hájek, P., and T. Hávranek (1978). *Mechanizing Hypothesis Formation*. New York: Springer-Verlag.

Hansel, C. E. M. (1980). *ESP and Parapsychology*. Buffalo, New York: Prometheus Books.

Hanson, N. R. (1958). *Patterns of Discovery*. Cambridge: Cambridge University Press.

Harnad, S. R., H. D. Steklis, and J. Lancaster, Eds. (1976). *Origins and Evolution of Language and Speech. Annals of the New York Academy of Sciences*, Vol. 280.

Hebb, D. O. (1949). *The Organization of Behavior.* New York: Wiley.

Hebb, D. O. (1951). The role of neurological ideas in psychology. *Journal of Personality* **20**: 39–55.

Hebb, D. O. (1966). *A Textbook of Psychology.* Philadelphia: W. B. Saunders.

Hebb, D. O., W. E. Lambert, and G. R. Tucker (1971). Language, thought and experience. *Modern Language Journal* **55**: 212–222.

Hebb, D. O. (1976). Physiological learning theory. *J. Abnormal Child Psychol.* **4**: 309–314.

Hebb, D. O. (1980). *Essay on Mind.* New York: L. Erlbaum Assoc.

Hécaen, H., and M. L. Albert (1978). *Human Neuropsychology.* New York: Wiley and Sons.

Held, R., and A. Hein (1963). Movement-produced stimulation in the development of visually guided behavior. *J. Compar Physiol. Psychol.* **56**: 872–876.

Hellman, R. H. G., Ed. (1980). *Nonlinear Dynamics. Annals New York Acad. Sci.*, v. 357.

Helmholtz, H. von (1873). *Popular Lectures on Scientific Subjects.* London: Longmans, Green and Co.

Helmholtz, H. von (1879). *Die Tatsachen in der Wahrnehmung.* Berlin: Hirschwald.

Hempel, C. G. (1962). Deductive-nomological *vs.* statistical explanation. In H. Feigl and G. Maxwell, Eds., *Minnesota Studies in the Philosophy of Science* III, pp. 98–169.

Hennessy, J. W., and S. Levine (1979). Stress, arousal, and the pituitary–adrenal system: a psychoendocrine hypothesis. *Progress in Psychobiology and Physiological Psychology* **8**: 133–178.

Hesse, M. (1966). *Models and Analogies in Science.* Notre Dame, Ind.: University of Notre Dame Press.

Hessen, B. (1931). The social and economic roots of Newton's "Principia". In *Science at the Cross Roads* pp. 149–212. London: Frank Cass, 1971.

Hilbert, D. (1918). Axiomatisches Denken, *Mathematische Annalen* **78**: 405–415.

Hinde, R. A. (1970). *Animal behavior. A Synthesis of Ethology and Comparative Physiology*, 2nd ed. New York: McGraw-Hill.

Hintikka, J. (1962). *Knowledge and Belief.* Ithaca, New York: Cornell University Press.

Hintikka, J. (1969). *Models for Modalities.* Dordrecht: Reidel.

Hoffman, W. C. (1966). The Lie Algebra of Visual Perception. *J. Math. Psychol.* **3**: 65–98.

Holton, G. (1973). *Thematic Origins of Scientific Thought.* Cambridge. Mass.: Harvard University Press.

Holton, G. (1978). *The Scientific Imagination. Case Studies.* Cambridge. Mass.: Harvard University Press.

Hubel, D. H., and T. N. Wiesel (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex *J Physiol.* **160**: 106–154.

Hubel, D. H., and T. H. Wiesel (1977). Functional architecture of macaque visual cortex. *Proc. Royal Soc. (London) B*, **198**: 1–59.

Hume, D. (1739). *A Treatise of Human Nature.* L. A. Selbidge, Ed. Oxford: Clarendon Press.

Hunter, W. S. (1929). The sensory control of the maze habit in the white rat. *J. Genetic Psychol* **36**: 505–537.

Inhelder, B., and J. Piaget (1958). *The Growth of Logical Thinking from Childhood to Adolescence.* New York: Basic Books.

James, W. (1890). *Principles of Psychology*, 2 volumes. Repr. New York: Dover, 1950.

James, W. (1907). *Pragmatism.* New York: Meridian, 1955.

Jantsch, E. (1972). *Technological Planning and Social Futures.* London: Cassell.

Jerison, H. J. (1973). *Evolution of the Brain and Intelligence.* New York: Academic Press.

Johnson-Laird, P. N., and P. C. Wason, Eds. (1977). *Thinking: Readings in Cognitive Science.* Cambridge: Cambridge University Press.

Jusczyk, P. W., and R. M. Klein, Eds. (1980). *The Nature of Thought. Essays in Honor of D. O. Hebb.* Hillsdale, N. J.: Lawrence Erlbaum.

Kandel, E. R., and J. H. Schwartz, Eds. (1981). *Principles of Neural Science.* New York: Elsevier/North Holland.

Kant, I. (1787). *Kritik der reinen Vernunft,* 2nd ed. R. Schmidt, Ed., Hamburg: Meiner, 1930.

Kapitza, P. L. (1979). Plasma and the controlled thermonuclear reaction. *Science* **205**: 959–964.

Karmiloff-Smith, A., and B. Inhelder (1974/5). "If you want to get ahead, get a theory". *Cognition* **3**: 195–212.

Katona, G. (1939). *Organizing and Memorizing.* New York: Columbia University Press.

Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money.* Collected Writings Vol. VII. London: Macmillan and Cambridge University Press 1973.

Kinsbourne, M. (1971). The minor cerebral hemisphere as a source of aphasic speech. *Arch. Neurol.* **25**: 302–306.

Krechevsky, I. (1932). "Hypotheses" versus "chance" in the pre-solution period in sensory discrimination learning. University of California Publications in Psychology, Vol. 6, No. 3.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Kuhn, T. S. (1974). Second thoughts on paradigms. In F. Suppe, Ed., *The Structure of Scientific Theories,* pp. 459–482. Urbana, Ill.: University of Illinois Press.

Kyburg, H. J., Jr. (1961). *Probability and Logic of Rational Belief.* Middleton, Conn.: Wesleyan University Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In Lakatos and Musgrave, Eds., pp. 91–195.

Lakatos, I. (1978). *Philosophical Papers,* 2 vols. J. Worrall and G. Currie, Eds., Cambridge: Cambridge University Press.

Lakatos, I., and A. Musgrave, Eds. (1970). *Criticism and the Growth of Knowledge.* Cambridge: University Press.

Lalande, André (1938). *Vocabulaire technique et critique de la philosophie,* 3 vols., 4th ed. Paris: Alcan.

Lambert, W. E. (1981). Bilingualism and language acquisition. *Annals of the New York Academy of Science* **379**: 9–22.

Lashley, K. (1929). *Brain Mechanisms and Intelligence.* Chicago: University of Chicago Press.

Latour, B., and S. Woolgar (1979). *Laboratory Life. The Social Construction of Scientific Facts.* Beverly Hills, Cal.: Sage Publications.

Lederberg, J., and E. A. Feigenbaum (1968). Mechanization of inductive inference in organic chemistry. In B. Kleinmutz, Ed. *Formal Representation of Human Judgment,* pp. 187–218. New York: John Wiley and Sons.

Lehrer, K. (1974). *Knowledge.* Oxford: Clarendon Press.

Leibniz, G. W. von (1703). *Nouveaux essais sur l'entendement humain.* Paris: Flammarion, s.d.

Lenin, V. I. (1908). *Materialism and Empirio-Criticism.* London: Lawrence and Wishart, 1950.

Lenzen, W. (1980). *Glauben, Wissen und Wahrscheinlichkeit.* Wien–New York: Springer-Verlag.

Leonardo da Vinci (1492 or 1515). Moral Precepts for the Student of Painting. In J. P. Richter, Ed., *The Notebooks of Leonardo da Vinci.* J. P. Richter, Ed., 2 vols. London: Sampson Low, Marston, Searle and Rivington, 1883. Repr. New York: Dover, 1970.

Leuba, C. (1955). Toward some integration of learning theories: the concept of optimal stimulation. *Psychol. Reports* **1**: 27–33.

Levine, M. (1966). Hypothesis behavior by humans during discrimination learning. *J. Exper. Psychology* **71**: 331–338.

Levine, M. (1974). *A Cognitive Theory of Learning.* Hillside, N. J.: L. Erlbaum Assocs.

Lewis, C. I. (1946). *An Analysis of Knowledge and Valuation.* La Salle, Ill.: Open Court.

Lewis, W. (1955). *The Theory of Economic growth.* London: Allen and Unwin.

Lieberman, P. (1975). *On the Origins of Language: An Introduction to the Evolution of Human Speech.* New York: Macmillan.

Lindauer, M. (1971). *Communication Among Social Bees.* Cambridge, Mass.: Harvard University Press.

Lloyd, G. E. R. (1966). *Polarity and Analogy.* Cambridge: Cambridge University Press.

Locke, J. (1689). *An Essay Concerning Human Understanding.* London: Routledge and Kegan Paul, s.d.

Loehlin, J. C., and R. C. Nichols (1976). *Heredity, Environment, and Personality.* Austin: University of Texas Press.

Lovejoy, A. O. (1929). *The Revolt Against Dualism.* La Salle, Ill.: Open Court.

Luneburg, R. K. (1947). *Mathematical Analysis of Binocular Vision.* Princeton: Princeton University Press.

Luria, A. R. (1961). *The Role of Speech in the Regulation of Normal and Abnormal Behavior.* Oxford: Pergamon Press.

Luria, A. R. (1976). *Cognitive Development. Its Cultural and Social Foundations.* Cambridge, Mass.: Harvard University Press.

Mach, E. (1905). *Erkenntnis und Irrtum.* Leipzig: Barth.

Mach, E. (1910). *Populär-wissenschaftliche Vorlesungen*, 4th ed. Leipzig: Barth.

Mackey, M. C., and L. Glass (1977). Oscillation and chaos in physiological control systems. *Science* **197**: 287–289.

Masterton, R. B., C. B. G. Campbell, M. E. Bitterman, and N. Hotton, Eds. (1976a). *Evolution of Brain and Behavior in Vertebrates.* Hillsdale, N. J.: L. Erlbaum Assoc.

Masterton, R. B., W. Hodos, and H. J. Jerison, Eds. (1976b). *Evolution, Brain, and Behavior.* Hillsdale, N. J.: L. Erlbaum Assoc.

Maudsley, H. (1876). *The Physiology of Mind.* London: Macmillan.

Mayer, R. E. (1977). *Thinking and Problem-Solving.* Glenview, Ill.: Scott, Foresman and Co.

McCloskey, M., A. Caramazza, and B. Green (1980). Curvilinear motion in the absence of forces: naive beliefs about the motion of objects. *Science* **210**: 1139–1141.

McKenna, M. C. (1976). Comments on Radinsky's "Later mammal radiations". In R. B. Masterton *et al.*, Eds. *Evolution of Brain and Behavior in Vertebrates.* Hillsdale, N. J.: Lawrence Erlbaum Associates.

Medawar, P. B. (1969). *Induction and Intuition in Scientific Thought.* Philadelphia: American Philosophical Society.

Meier, N. R. F. (1931). Reasoning in humans. II. *J. Comparative Psychology* **12**: 181–194.

Mermin, N. D. (1981). Quantum mysteries for anyone. *J. Phil.* **78**: 397–408.

Merton, K. (1973). *The Sociology of Knowledge.* Chicago: University of Chicago Press.

Mill, J. S. (1875). *A System of Logic*, 8th ed. London: Longmans, Green, 1952.

Miller, G. A. (1967). *The Psychology of Communication.* New York: Basic Books.

Milner, P. (1970). *Physiological Psychology.* New York: Holt, Rinehart and Winston.

Minsky, M. (1977). Frame-system theory. In Johnson-Laird and Wason, Eds. pp. 355–376.

Modigliani, F. (1977). The monetarist controversy or, should we forsake stabilization policies? *Amer. Econ. Rev.* **67**: 1–19.

Morton, J., and J. C. Marshall, Eds. (1977). *Psycholinguistics Series 1: Developmental and Pathological.* London: Elk Science.

Moruzzi, G. and Magoun, H. W. (1949). Brain stem reticular formation and activation of the EEG. *Electroencephalography and Clinical Neurophysiology* **1**: 455–473.

Myrdal, G. (1969). *Objectivity in Social Research.* New York: Pantheon Books.

Myrdal, G. (1973). *Against the Stream.* New York: Pantheon Books.

Nagel, E. (1961). *The Structure of Science.* New York: Harcourt, Brace and World.

Neisser, U. (1963). The multiplicity of thought. *Brit. J. Psychol.* **54**: 1–14.

Neisser, U. (1976). *Cognition and Reality.* San Francisco: Freeman.

Neisser, U. (1980). *The Limits of Cognition.* In Jusczyk and Klein, Eds., pp. 115–132.

Nelson, L. (1912). The impossibility of the "theory of knowledge". In *Socratic Method and Critical Philosophy: Selected Essays.* Transl. T. K. Brown III. New Haven, Conn.: Yale University Press. Repr.: New York, Dover, 1965.

Newell, A., and H. Simon (1972). *Human Problem Solving.* Englewood Cliffs, N. J.: Prentice-Hall.

Nisbett, R., and L. Ross (1980). *Human Inference: Strategies and Shortcomings of Social Judgment.* Englewood Cliffs, N. J.: Prentice-Hall.

Noton, D., and L. Stark (1971). Scanpaths in the eye movements during pattern perception. *Science* **171**: 308–311.

Nottebohm, F. (1981). A brain for all seasons: cyclical anatomical changes in song control nuclei of the canary brain. *Science* **214**: 1368–1370.

Oakeshott, M. (1962). *Rationalism in Politics and other Essays.* London: Methuen.

O'Keefe, J., and L. Nadel (1978). *The Hippocampus as a Cognitive Map.* Oxford. Clarendon Press.

Olds, J. (1975). Mapping the mind onto the brain. In F. G. Worden, J. P. Swazey, and G. Adelman, Eds. (1975). *The Neurosciences: Paths of Discovery.* Cambridge, Mass.: MIT Press.

Patterson, F. (1978). Linguistic capabilities of a low-land gorilla. In F. C. C. Peng, Ed., *Sign Language Acquisition in Man and Ape.* Boulder Colo.: Westview Press.

Patterson, F. (1981). Ape language. *Science* **211**: 86–87.

Pavlov, I. P. (1927). *Conditioned Reflexes.* New York: Dover, 1960.

Pavlov, I. P. (1955). *Selected Works.* Moscow: Foreign Languages Publ. House.

Penfield, W., and T. Rasmussen (1950). *The Cerebral Cortex of Man.* New York: Macmillan.

Phelps, M., D. E. Kuhl, and J. C. Mazziotta (1981). Metabolic mapping of the brain's response to visual stimulation: studies in humans. *Science* **211**: 1445–1448.

Piaget, J. (1952). *The Origins of Intelligence in Children.* New York: International Universities Press.

Piaget, J. (1955). *The Language and Thought of the Child.* New York: Meridian.

Piaget, J. (1967). *Biologie et connaissance.* Paris: Gallimard.

Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. *Human Development* **15**: 1–12.

Piaget, J. (1976a). *The Psychology of Intelligence.* Totowa, N. J.: Adams.

Piaget, J. (1976b). *Le comportement, moteur de l'évolution*. Paris: Gallimard.

Piaget, J., and B. Inhelder (1967). *The Child's Conception of Space*. New York: Norton.

Pincus, J. H., and G. J. Tucker (1978). *Behavioral Neurology*, 2nd ed. New York: Oxford University Press.

Polanyi, M. (1958). *Personal Knowledge*. Chicago, University of Chicago Press.

Pólya, G. (1954). *Mathematics and Plausible Reasoning*, 2 vols. Princeton: Princeton University Press.

Pólya, G. (1957). *How to Solve it*? New York: Doubleday Anchor Books.

Popper, K. R. (1957). *The Poverty of Historicism*. London: Routledge and Kegan Paul.

Popper, K. R. (1959). *The Logic of Scientific Discovery*. London: Hutchinson.

Popper, K. R. (1963). *Conjectures and Refutations*. New York: Basic Books.

Popper, K. R. (1972). *Objective Knowledge*. Oxford: Clarendon Press.

Popper, K. R. (1974). Intellectual Autobiography. In P. Schilpp, Ed., *The Philosophy of Karl R. Popper*.

Popper, K. R., and J. C. Eccles (1977). *The Self and its Brain*. New York: Springer-Verlag.

Premack, D. (1971). Language in chimpanzee? *Science* **172**: 808–822.

Putnam, H. (1960). Minds and machines. In S. Hook, Ed., *Dimensions of Mind*, pp. 148–179. New York: New York University Press.

Putnam, H. (1975). *Philosophical Papers*, 2 vols. Cambridge: Cambridge University Press.

Pylyshyn, Z. W. (1980). Computation and cognition: issues in the foundations of cognitive science, *The Behavioral and Brain Sciences* **1**: 111–132.

Ramsey, F. P. (1931). *The Foundations of Mathematics*. London: Routledge and Kegan Paul.

Randi, J. (1979). A controlled test of dowsing abilities. *The Skeptical Inquirer* **4**: 16–29.

Rasiowa, H., and R. Sikorski (1963). *The Mathematics of Metamathematics*. Warsaw: Panstwowe Wydawnictwo Naukowe.

Ratliff, F. (1965). *Mach Bands: Quantitative Studies on Neural Networks of the Retina*. S. Francisco: Holden-Day.

Ratliff, F. (1971). Illusions in man and his instruments. *J. Phil.* **LXVIII**: 591–597.

Rescher, N. (1976). *Plausible Reasoning*. Assen-Amsterdam: Van Gorcum.

Rescher, N. (1977). *Methodological Pragmatism*. New York: New York University Press.

Restle, F. (1976). The selection of strategies in cue learning. *Psychol. Rev.* **69**: 329–343.

Rignano, E. (1923). *The Psychology of Reasoning*. Transl. W. A. Holl London: Kegan Paul, Trench, Trubner and Co.

Roche, M. (1976). Factors governing the scientific and technological development of a country. *Scientia* **111**: 75–84.

Rorty, R. (1979). *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.

Rosch, E. (1974). Linguistic relativity. In A. Silverstein, Ed., *Human Communication: theoretical Perspectives*. Hillsdale, N. J.: L. Erlbaum Assoc.

Russell, B. (1914). *Our Knowledge of the External World*. London: George Allen and Unwin.

Russell, B. (1918). *Mysticism and Logic, and Other Essays*. London: Longmans, Green.

Russell, B. (1940). *An Inquiry into Meaning and Truth*. London: George Allen and Unwin.

Russell, B. (1948). *Human Knowledge. Its Scope and Limits*. London: Allen and Unwin.

Ryle, G. (1954). *Dilemmas*. Cambridge University Press.

Sampson, G. (1978). Linguistic universals as evidence for empiricism. *J. Linguistics* **14**: 183–206.

Sánchez, F. (1581). *Que nada se sabe [De multum nobili et prima universali scientia quod nihil scitur]*. Buenos Aires: Emecé, s.d.

Savage, L. J. (1972). *The Foundations of Statistics*, 2nd ed. New York: Dover.

Savage-Rumbaugh, E. S., D. M. Rumbaugh, and S. Boysen (1978). Symbolic communication between two chimpanzees. *Science* **201**: 641–644.

Schilpp. P. A., Ed. (1974). *The Philosophy of Karl R. Popper*, 2 vols. La Salle, Ill.: Open Court.

Schlick, M. (1925). *General Theory of Knowledge*. Transl. A. E. Blumberg. New York–Wien: Springer-Verlag, 1974.

Scribner, S. (1977). Modes of thinking and ways of speaking: culture and logic reconsidered. In Johnson-Laird and Wason, Eds. pp. 483–500.

Sebeok, T. A., and R. Rosenthal, Eds. (1981). *The Clever Hans Phenomenon: Communication With Horses, Whales, Apes, and People*. New York: New York Academy of Sciences.

Selfridge, O. G., and U. Neisser (1960). Pattern recognition by machine. *Sci. Amer.* **203** (Aug.): 60–68.

Shannon, C., and W. Weaver (1949). *The Mathematical Theory of Communication*. Urbana, Ill.: University of Illinois Press, 1963.

Sherif, M. (1936). *The Psychology of Social Norms*. New York: Harper and Bros.

Sherratt, A. (Ed.) (1980). *The Cambridge Encyclopedia of Archaeology*. New York: Cambridge University Press.

Shimony, A. (1971). Perception from an evolutionary point of view. *J. Phil.* **68**: 571–583.

Simon, H. A., and G. J. Groen (1973). Ramsey elimination and the testability of scientific theories. *Brit. J. Phil. Sci.* **24**: 367–380.

Simon, H. A. (1979). Information processing models of cognition. *Annual Rev. Psychol.* **30**: 363–396.

Skinner, B. F. (1948). Superstition in the pigeon. *J. Exp. Psychol.* **38**: 168–172.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson and D. I. Slobin, Eds., *Studies of Child Language Development*, New York: Holt, Rinehart and Winston.

Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian. J. Psychol.* **4**: 165–173.

Smith-Churchland, P. (1978). Fodor on language learning. *Synthese* **38**: 149–159.

Smith-Churchland, P. (1980). A perspective on mind–brain research. *J. Phil.* **LXXVII**: 185–207.

Sneed, J. D. (1979). *The Logical Structure of Mathematical Physics*, 2nd ed. Dordrecht: Reidel.

Stegmüller, W. (1976). *The Structure and Dynamics of Theories*. New York: Springer-Verlag.

Stich, S. P. (1979). Between Chomskyan rationalism and Popperian empiricism. *Brit. J. Phil. Sci.* **30**: 329–347.

Suppe, F., Ed. (1974). *The Structure of Scientific Theories*. Urbana, Ill.: University of Illinois Press.

Suppes, P. (1970). *A Probabilistic Theory of Causality*. Acta Philosophica Fennica XXIV. Amsterdam: North-Holland.

Suppes, P., Ed. (1978). *Impact of Research on Education: Some Case Studies*. Washington, D.C.: National Academy of Sciences.

Swets, J. A. (1973). The relative operating characteristics in psychology. *Science* **182**: 990–1000.

Szentagothai, J. (1978). The neuron network of the cerebral cortex. *Proc. Roy. Soc. London* **B201**: 219–248.

Tarski, A. (1953). A general method in proofs of undecidability. In A. Tarski, A. Mostowski, and R. M. Robinson, *Undecidable Theories*. Amsterdam: North-Holland.

Tarski, A. (1956). *Logic, Semantics, Metamathematics*. Oxford: Clarendon Press.

Taylor, J. G. (1962). *The Behavioral Basis of Perception*. New Haven: Yale University Press.

Terrace, H. S., L. A. Pettito, R. J. Sanders, and T. G. Beaver (1979). Can an ape create a sentence? *Science* **206**: 891–909.

Teuber, H. L. (1965). Convergences, divergences, lacunae. In J. C. Eccles, Ed., *Brain and Conscious Experience*, pp. 575–583. New York: Springer-Verlag.

Thomas, R. D. K., and E. C. Olson (1980). *A Cold Look at Warm-Blooded Dinosaurs*. Boulder, Colorado: Westview Press.

Thompson, R. F. (1975). *Introduction to Physiological Psychology*. New York: Harper and Row.

Thompson, R. F., M. M. Patterson, and T. J. Teyler (1972). The neurophysiology of learning. *Annual Rev. Psychol.* **23**: 73–104.

Thurstone, L. L. (1930). The learning function. *J. General Psychology* **3**: 469–491.

Tolman, E. C., and I. Krechevsky (1933). Means-end-readiness and hypothesis. *Pychological Review* **40**: 60–70.

Törnebohm, H. (1979). Paradigms in Fields of Research. In I. Niiniluoto and R. Tuomela, Eds., *The Logic and Epistemology of Scientific Change*, pp. 62–90. Amsterdam: North-Holland.

Trehub, A. (1977). Neuronal models for cognitive processes: networks for learning, perception and imagination. *J. Theoret. Biology* **65**: 141–169.

Trudgill, Peter (1974). *Sociolinguistics*. Harmondsworth, Penguin Books.

Truesdell, C. (1974). A Simple example of an initial-value problem with any desired number of solutions. *Rendiconti dell'Istituto Lombardo*, Classe di Scienze (A) **108**: 301–304.

Truesdell, C. (1981). The role of mathematics in science as exemplified by the work of the Bernoullis and Euler. *Verhandl. Naturf. Ges. Basel* **91**: 5–22.

Truesdell, C. (1983). *An Idiot's Fugitive Essays on Science*. New York: Springer-Verlag.

Tuchman, B. (1978). *A Distant Mirror*. New York: Ballantine Books.

Tuomela, R. (1973). *Theoretical Concepts*. Wien-New York: Springer-Verlag.

Uexküll, J. von (1921). *Umwelt und Innenwelt der Tiere*, 2nd ed. Berlin: Springer-Verlag.

Umiker-Sebeok, J., and T. A. Sebeok (1981). Clever Hans and smart simians. *Anthropos* **76**: 89–165.

Unger, P. (1975). *Ignorance: A Case for Scepticism*. Oxford: Clarendon Press.

Vaihinger, H. (1920). *Die Philosophie des Als Ob*, 4th ed. Leipzig: Meiner.

Ville, J. (1939). *Etude critique de la notion de collectif*. Paris: Gauthier-Villars.

Vogt, E. Z., and R. Hyman (1959). *Water Witching USA*. Chicago: University of Chicago Press.

Volkmann, F. C., L. A. Riggs, and R. K. Moore (1980). Eyeblinks and visual suppression. *Science* **207**: 900–902.

Vollmer, G. (1975). *Evolutionäre Erkenntnistheorie*. Frankfurt: S. Hirzel.

Vygotskii, L. S. (1962). *Thought and Language*. Transl. E. Hanfman and G. Vakar. Cambridge, Mass.: MIT Press.

Walcott, C., J. L., Gould, and J. L. Kirschvink (1979). Pigeons have magnets. *Science* **205**: 1027–1029.

Ward, W. C., and H. M. Jenkins (1965). The display of information and the judgment of contingency. *Can. J. Psych.* **19**: 231–241.

Warnock, G. J. (1958). *English Philosophy Since 1900*. London: Oxford University Press.

Wason, P. C., and P. N. Johnson-Laird (1972). *Psychology of Reasoning*. London: Batsford.

Wasserman, G. S., G. Felstem, and G. E. Easland (1979). The psychophysical function: harmonizing Fechner and Stevens. *Science* **204**: 85–87.

Weiskrantz, L., and E. K. Warrington (1979). Conditioning in amnesia patients. *Neuropychologia* **17**: 187–193.

Wertheimer, M. (1961). *Productive Thinking*. London: Tavistock.

Westfall, R. S. (1973). Newton and the fudge factor. *Science* **179**: 751–758.

Whewell, W. (1847). *The Philosophy of the Inductive Sciences*, 2 vols. London: Frank Cass, 1967.

Whorf, B. L. (1956). *Language, Thought and Reality*. New York: Wiley.

Wiesel, T. N., and D. H. Hubel (1965). Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. *J. Neurophysiol.* **28**: 1029–1040.

Wilder, R. (1981). *Mathematics as a Cultural System*. Oxford: Pergamon.

Williams, B. A. O. (1972). Knowledge and reasons. In G. H. v. Wright, Ed., *Problems in the Theory of Knowledge*. The Hague: Martinus Nijhoff.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. London: Routledge and Kegan Paul.

Wittgenstein, L. (1953). *Philosophical Investigations*. New York: Macmillan.

Wolpe, J. (1976). How laboratory-derived principles of learning have conquered the neuroses. In G. Serban, Ed. *Psychopathology of Human Adaptation*. New York: Plenum.

Wolpe, J. (1978). Cognition and causation in human behavior and its therapy. *Amer. Psychologist* **33**: 437–446.

Woodruff, G., D. Premack, and K. Kennel (1978). Conservation of liquid and solid quantity by the chimpanzee. *Science* **202**: 991–994.

Woodruff, G., and D. Premack (1981). Primative mathematical concepts in the chimpanzee: proportionality and numerosity. *Nature* **293**: 568–570.

Ziman, J. (1968). *Public Knowledge*. Cambridge: Cambridge University Press.

Ziman, J. (1979). *Reliable Knowledge*. New York: Cambridge University Press.

INDEX OF NAMES

INDEX OF SUBJECTS